Epigenetic Profiling and Molecular Characterisation of Non-melanoma Skin Cancer

Mladkova, Nikol

Epigenetic Profiling and Molecular Characterisation of Non-melanoma Skin Cancer

Nikol Mladkova, MD, MPH

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Centre for Cutaneous Research,
The Blizard Institute,
Barts and The London School of Medicine and Dentistry,
Queen Mary, University of London

Supervisors:
Professor Catherine Harwood
Professor Vardhman Rakyan
Professor Charlotte Proby

August 2014

I, Nikol Mladkova, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature: Nikol Mladkova

Date: 26th August, 2014

Contributions

I am grateful to the following persons for their contributions to work presented in this thesis:

Dr Cleo Bishop (Queen Mary, University of London) carried out the culture and RNA isolation of embryonic stem cells.
Dr Thomas Down (University of Cambridge) analysed whole-genome bisulfite-sequencing data, and contributed to the analysis of Illumina 450K array data.
Angela McHugh (University of Dundee) carried out the culture and DNA isolation of cell lines hybridised to Illumina 27K array.

Abstract

Non-melanoma skin (NMSC) cancer is the most common human malignancy. Cutaneous squamous cell carcinoma (cSCC) and its precursor, actinic keratosis (AK) affect tens of thousands of people each year in the UK. Merkel cell carcinoma is a rare, yet aggressive type of NMSC recently linked with Merkel Cell Polyomavirus (MCPyV). In spite of the clinical burden of NMSC, key molecular regulatory patterns remain largely unknown. The aims of this thesis were to investigate genome-wide genetic, epigenetic and transcriptional changes in AK and cSCC, and assess the prevalence of MCPyV and its effect on methylation in NMSC.

Copy-number analysis revealed that AK harbours significantly more genomic aberrations compared to skin, the majority of which occurs on chromosomes 8 and 9. Transcriptional profiling has found 292 and 308 genes as differentially expressed in AK compared to non-sunexposed and sun-exposed skin, respectively, and gene-set enrichment analysis (GSEA) revealed dysregulation of PPAR pathway in this lesion.

Expression profiling of cSCC and AK has revealed 346 differentially expressed genes, and GSEA detected dysregulation in several canonical pathways including TGF-β and MAPK pathway. Aberrant methylation in cSCC cell lines occurs in the promoters of many developmental genes. A total of 1085 hyper- and 833 hypomethylated genes were detected in cSCCs, and GSEA revealed dysregulation of critical signalling pathways (WNT, MAPK signalling pathways). Methylation analysis of AK revealed a total of 4194 differentially methylated genes, and implicated FOXF2, PITX2, RUNX1 and SMAD3 transcription factors in this lesions.

MiRNA profiling of cSCC and normal skin revealed significant dysregulation of 38 miRNAs including several of viral origin.

MCPyV was shown to be common in NMSC, yet MCPyV nor human papillomavirus does not affect cSCC methylation.

Taken together, this work provides novel insight into molecular regulation of cSCC oncogenesis, and identifies potential epigenetic targets for functional evaluation in this malignancy.

Acknowledgments

I feel eternally indebted to my supervisors, Professors Catherine Harwood, Vardhman Rakyan and Charlotte Proby, for allowing me the privilege of working on this wonderful skin cancer project, and thus fulfilling my dream of becoming a cancer researcher. For that I shall forever be grateful.

Specifically, I could never thank enough Professor Harwood for her inexhaustible enthusiasm, generous support, and for allowing me to spread my wings. Professor Rakyan has not only been incredibly kind, honest and fair to me, but has also been an endless source of inspiration and a true role model as a scientist. I consider it an honour to know him and having him as my supervisor. Professor Proby deserves my gratitude for her firm support, élan vital and for ensuring we keep our feet on the ground.

This project would not have been possible without the generous support of the British Skin Foundation that provided my studentship, and the Barts and the London Charity research grant. I sincerely thank both organisations for providing funds that allowed this project to happen.

I cannot underscore enough how grateful I am to Dr Thomas Down not only for his collaboration on the bioinformatic analysis, but namely for his inspiring attitude that motivated me to develop and pursue my own biostatistics/bioinformatics learning early on in the process. He uncovered a whole new dimension of science for me, which I very much enjoy.

Professors Denise Sheer and Dean Nizetic have both provided invaluable advice and enthusiastic support as my scientific committee, and I can only secretly hope to work with people of such sterling integrity and intellectual breadth in the future.

I am very grateful to Dr Cory Adamson of Duke University, who has started me on the path of cancer research many years ago in a way that nobody else could and encouraged me all along the road, to Dr Richard Mark White and Professor Leonard Zon of Harvard University for allowing me to join the laboratory to pick up skill I did not know existed. I had a truly wonderful time, and want to thank you both for it.

All patients who agreed to participate in our research deserve nothing but my deepest respect and my endless gratitude. They are the very real faces and stories behind genes, protein binding partners and cellular regulation processes and may I never forget that.

I am also grateful to Drs Sacha Goolamali and Tang Shim for being cool and funny lab mates and very apt pupils I had a supreme pleasure teaching supremely boring techniques, to the Rakyan gang for being sweet and unprecedentedly friendly colleagues. Namely Carolina Gemma, who is a great teacher, Michelle Holland, Guillermo Carbajosa, Alex Bazeos, and Melissa Smart.

I would like to extend my thanks to my peers and friends, namely Folake, Adiam, Anissa and Charlotte for all the good coffees and good moans, served as necessary.

I also appreciate many useful cell culture tips I have received from Sree and Dr John Connelly.

A special mention goes to the laboratory manager Jeff Maskell for keeping me out of harm's way (even if against my will sometimes), to Ray Grundwell of the digital and photographic imaging centre for preparing my many conference posters perfectly and always on time; Dr Emily Clemente of Cambridge University and her team for processing my hard-mined DNA and RNA samples diligently and with care, to Gary Warnes for sorting out my FACS sorting, and to laboratory assistants and all helpful staff that provided clean and safe environment without which my research could not have happened, yet never demand nor receive any credit for it. They deserve it all the more.

I wish to express my most sincere gratitude to my parents Rudolf and Danuše, to my sister Barbara, for never vacillating in their support of my research work, even though it involved my transfer to a foreign country for many years and many sleepless nights. I could have never done this without them. I also wish to thank my husband Jakub not only for his support and care in difficult times, but also for his unquenching faith in me in moments when I have lost all of mine.

Finally, it remains to thank Mother Nature for being a continuing source of fascination for me, and may I never become a hollow specter filled with fatuous conceits in my pursuit of her great and ancient secrets.

Dedication

To my grandparents Jozefina Sekerková (1930-2010), Isidor Sekerka (1931-2009), and to my aunt Anna Novotná née Sekerková (1956-2011), who have departed my life during work on this PhD thesis.

My grandparents have taught me respect for hard work, for all living creatures, and to believe in my dreams. They are a big part of who I am.

My aunt, who has departed my life a way too prematurely, has - perhaps unconsciously - encouraged me to pursue my dreams and destiny while she herself was never allowed this privilege, and to face life's challenges proudly and with a bright smile, no matter how insurmountable those seem.

I miss you all every day.

To my daughter Shirin Anna, who has arrived in my life during work on this PhD thesis.

I love you endlessly.

List of chapters

3. Somatic Copy Number Alterations and Gene Transcription Profiling of Normal Skin and Actinic Keratoses 140

9. Summary of results and concluding remarks 403

Appendix 1. Differentially expressed genes between SE skin and AK. 409

List of figures

List of tables

Abbreviations

Abbreviation	Description
°C	Degree Celsius
AIDS	Acquired immune deficiency syndrome
AK	Actinic keratosis
Akt	Protein kinase B
BCC	Basal cell carcinoma
BCL2	B-cell lymphoma 2
bp	base pairs
BRAF	B-Raf proto-oncogene
BS-seq	Bisulfite sequencing
CDK	Cyclin-dependent kinase
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	Complementary DNA
CGH	Comparative genomic hybridisation
ChIP	Chromatin immunoprecipitation
CIP/KIP	CDK interacting protein/Kinase inhibitory protein
CLL	Chronic lymphocytic leukaemia
CML	Chronic myelogenous leukemia
CNV	Copy-number variation
CO2	Carbon dioxide
CpG	a cytosine-phosphate-guanine sequence, double stranded
CSC	Cancer stem cells
cSCC	Cutaneous Squamous Cell Carcinoma
Cy3	Cyanine 3
Cy5	Cyanine 5
DAPI	4',6-diamidino-2-phenylindole
DEG	Differentially expressed genes
DMEM	Dulbecco’s Modified Eagle’s Medium

DMG	Differentially methylated gene
DMP	Differentially methylated probe
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA	Deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyl transferase
DNMT1	DNA methyltransferase 1
dNTP	Deoxynucleotide Triphosphate
dsDNA	Double-stranded DNA
EBV	Epstein–Barr virus
ECM	Extra Cellular Matrix
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-mesenchymal transition
ESC	Embryonic stem cells
FACS	Fluorescence-activated cell sorting
FAP	Familial adenomatous polyposis
FC	Fold-change
FCS	Fetal calf serum
FFPE	Formalin-fixed paraffin embedded
FHIT	Fragile histidine triad protein
FISH	Fluorescence in situ hybridisation
FITC	Fluorescein isothiocyanate
FOXN1	Forkhead box protein N1
g	gram
g	gravity acceleration
GFR	Growth factor receptors
GO	Gene ontology
GSEA	Gene set enrichment analysis

GWAS	Genome-wide association studies
H&E	Hematoxilin and eosin
H3K27me3	histone H3 lysine 27 trimethylation
H3K36me3	histone H3 lysine 36 trimethylation
H3K4me3	histone H3 lysine 4 trimethylation
HBV	hepatitis B virus
HDAC	Histone deacetylase
hESC	Human embryonic stem cells
HHV-8	Human herpesvirus 8
HIV	Human immunodeficiency virus
HP1A	Heterochromatin protein 1A
HPV	Human papillomavirus
HSC	Hematopoietic Stem Cell
HTLV-1	Human T-lymphotropic virus Type I
HTLV-2	The Human T-lymphotropic virus Type II
IC	Immunocompetent
ICM	Inner cell mass
IPSC	Induced pluripotent stem cells
IS	Immunosuppressed
JSRV	Jaagsiekte sheep retrovirus
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIT	c-Kit
KRT	Keratin
LCM	Laser capture microdissection
LCR	Long control region
LOH	Loss of heterozygosity
M	moles
MALT	Mucosa-associated lymphoid tissue
MAPK	Mitogen activated protein kinase

MBD	methyl-CpG binding domain
MCC	Merkel Cell Carcinoma
MCPyV	Merkel Cell Polyoma virus
MD	Moderately differentiated
MD-PD	Moderately-to-poorly differentiated
MEF	Mouse embryonic fibroblasts
MET	MET proto-oncogene, receptor tyrosine kinase
mg	milligram
MgCl2	magnesium chloride
min	minute
miRNA	MicroRNA
Mm	Milimeter
mRNA	messenger RNA
MSC	Mesenchymal stem cells
MsigDB	Molecular Signatures Database
mTOR	Mammalian Target of Rapamycin
MYC	c-Myc
ng	nanogram
NHK	Normal human keratinocytes
NHS	National healthcare system
nm	nanometer
NMSC	Non-melanoma skin cancer
NSCLC	Non-small cell lung cancer
OCT	optimal cutting temperature compound
OTR	Organ transplant recipient
p14arf	Alternate reading frame product of cyclin-dependent kinase inhibitor 2A
p16INK4a	Cyclin-dependent kinase inhibitor 2A
P21	Cyclin-dependent kinase inhibitor 1
P27	Cyclin-dependent kinase inhibitor 1B

P53	Tumour protein 53
P57	Cyclin-dependent kinase inhibitor 1C
PBS	phosphate buffered saline
PCR	Polymerase chain reaction
PD	Poorly differentiated
PE	Phycoerythrin
PEN	polyethylene napthalate
PHK	Primary human keratinocytes
pRb	Retinoblastoma protein
Pri-miRNA	primary miRNA
PTCH	Patched
PTEN	Phosphatase and tensin homolog
PTPRD	protein tyrosine phosphatase receptor type d
QPCR	Quantitative PCR
RAF	c-Raf
RB	Retinoblastoma gene
RDEB	Recessive Dystrophic Epidermolysis Bullosa
RISC	RNA-induced silencing complex
RKIP	Raf kinase inhibitor protein
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
RSV	Rous sarcoma virus
RTR	Renal transplant recipient
SCC	Squamous cell carcinoma
SCNA	Somatic copy number alteration
SCNT	Somatic-cell nuclear transfer
SD	standard deviation
SEM	standard error of the mean

SNP	single nucleotide polymorphism
SV40	Simian vacuolating virus 40
TBE	Tris Borate EDTA
TET	ten-eleven translocation protein
TF	transcription factor
Tm	Melting temperature
TP63	Tumour protein 63
UPD	Uniparental disomy
UTR	Untranslated region
UV	Ultraviolet
UVB	Ultraviolet-B
UVC	Ultraviolet-C
UVR	Ultraviolet radiation
V	Volt
WD	Well differentiated
WMD	Well-to-moderately differentiated
ΔNp63	amino-deleted p63 isoforms
μg	Microgram

Publications, Presentations and Awards

Publications: Journal Articles

Wang H, Lee S, Nigro CL, Lattanzio L, Merlano M, Monteverde M, Matin R, Puride K, **Mladkova** N, et al. NT5E (CD73) is epigenetically regulated in malignant melanoma and associated with metastatic site specificity. British journal of cancer. 2012;106(8):1446-52.

Lambert S, **Mladkova** N, Gulati A, Hamoudi R, Purdie K, Cerio R, et al. Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome profiling. British journal of cancer. 2014;110(2):520-9.

Mladkova, N, McHugh, A, Down, T, Rakyan, VK, Proby, CM, Harwood, CA. Identification of Differentially Methylated Genes in Cutaneous Squamous Cell Carcinoma (manuscript in preparation).

Mladkova, N, Bishop, C, Rakyan, VK, Tsai, MS, Hwang, SM, Proby, CM, Harwood, CA. Gene Expression Profiling of Cutaneous Squamous Cell Carcinoma Reveals Convergence Toward a Mesenchymal Stem Cells Transcriptome in Less Differentiated Histological Tumour Subtypes (manuscript in preparation).

Publications: Abstracts

- Lambert, S. R., **Mladkova, N**., Gulati, A., Leigh, I. M., Harwood, C. A., & Proby, C. M. (2011, April). Microarray expression profiling identifies novel targets and pathways involved in cutaneous squamous cell carcinoma oncogenesis. In BRITISH JOURNAL OF DERMATOLOGY (Vol. 164, No. 4, pp. 933-934).

- **Mladkova, N.**, Kader, P., Harwood, C., & Proby, C. (2011, April). Are immunosuppressed patients at greater risk of Merkel cell polyomavirus in non-melanoma skin cancer?. In BRITISH JOURNAL OF DERMATOLOGY (Vol. 164, No. 4, pp. 937-937).

- **Mladkova** N, McHugh A, Down T, Rakyan V, Proby C, Harwood C. Identification of Differentially Methylated Genes in Cutaneous Squamous Cell Carcinoma from Epidermolysis Bullosa Patients and Normal Population. In JOURNAL OF INVESTIGATIVE DERMATOLOGY (Vol. 131, pp. S24-S24).

- **Mladkova, N.,** McHugh, A., Down, T., Rakyan, V., South, A., Proby, C., & Harwood, C. (2011, April). Methylation profiling of cutaneous squamous cell carcinoma. In BRITISH JOURNAL OF DERMATOLOGY (Vol. 164, No. 4, pp. 932-933).

- **Mladkova, N.,** McHugh, A., Rakyan, V., Proby, C., & Harwood, C. (2012, April). Methylation profiling of cutaneous squamous cell carcinoma reveals an embryonic stem cell signature among differentially methylated genes. In BRITISH JOURNAL OF DERMATOLOGY (Vol. 166, No. 4, pp. e23-e23).

- **Mladkova, N**., Cerio, R., Proby, C., & Harwood, C. (2011, June). Merkel cell carcinoma: the possible pathogenic role of viruses and the dilemma of managing metastatic skin malignancy in the context of immunosuppression. In BRITISH JOURNAL OF DERMATOLOGY (Vol. 164, No. 6, pp. 1410-1410).

- Lambert, S. R., **Mladkova, N**., Gulati, A., Hamoudi, R., Purdie, K. J., Cerio, R., ... & Harwood, C. A. (2013, May). Expression microarray analysis identifies key processes in the transition from actinic keratosis to cutaneous squamous cell carcinoma. In JOURNAL OF INVESTIGATIVE DERMATOLOGY (Vol. 133, pp. S67-S67).

- Goolamali, S. J., **Mladkova, N.**, Purdie, K. J., Shim, T. N., de Koning, M. N., Quint, W. G. & Bunker, C. B. (2013, May). HPV in non-melanoma skin cancer/pre-cancer in HIV. In JOURNAL OF INVESTIGATIVE DERMATOLOGY (Vol. 133, pp. S173-S173).

- **Mladkova, N.**, Rakyan, V., Proby, C., & Harwood, C. (2014, August). Genome-wide methylation profiling of cutaneous squamous cell carcinoma reveals novel epigenetically regulated genes and pathways. In JOURNAL OF INVESTIGATIVE DERMATOLOGY (Vol. 134, No. 8, pp. S6-S6).

Presentations

Location	Date and presentation title
William Harvey Day, 2009	20th October, 2009 Poster presentation – Methylation profiling of cutaneous squamous cell carcinoma
William Harvey Day, 2010	26th October, 2010 Poster presentation – Expression profiling of cutaneous squamous cell carcinoma
Postgraduate Research Day, 2011	April, 2011 Poster presentation – The Role of Merkel Cell Polyoma Virus in non-melanoma skin cancer
British Association of Dermatology	24th March, 2011 Oral presentation - Merkel Cell Carcinoma in a Renal Transplant Patient
3rd HIV symposium	May, 2011 Poster presentation – The role of HPV virus in non-melanoma skin cancer of HIV patients
ISCOMS, 2010 **Groningen, The Netherlands**	8-11th of June, 2010 Oral presentation – Epigenetic regulation of cutaneous squamous cell carcinoma
SCOPE 10th Annual Meeting **Oslo, Norway**	June, 2010 Oral presentation – The role of Merkel Cell Polyomavirus in non-melanoma skin cancer Oral presentation – Expression Profiling of Cutaneous Squamous Cell Carcinoma
National Cancer Research Institute, 2010 **Liverpool, UK**	November 7-11, 2010 Poster presentation - Expression Profiling of Cutaneous Squamous Cell Carcinoma Poster presentation - The role of Merkel Cell Polyomavirus in non-melanoma skin cancer
British Society of Investigative Dermatology Meeting, 2010 **Manchester, UK**	April, 2011 Poster presentation - Expression Profiling of Cutaneous Squamous Cell Carcinoma Poster presentation - The role of Merkel Cell Polyomavirus in non-melanoma skin cancer
EMBO Conference Series **Chromatin and Epigenetics** **EMBL, Heidelberg, Germany**	May, 2011 Poster presentation – Epigenetic Profiling of Cutaneous Squamous Cell Carcinoma
National Cancer Research Institute, 2011	November 11-15, 2011 Poster presentation – Identification of Tumour-Inducing Cells in

Liverpool, UK	Squamous Cell Carcinoma Poster presentation – Identification of Differentially Methylated Genes in Cutaneous Squamous Cell Carcinoma from Epidermolysis Bullosa Patients and Normal Population
Society for Investigative Dermatology, 73rd Annual Meeting Albuquerque, NM, USA	May 7-10, 2014 Poster presentation - Genome-wide methylation profiling of cutaneous squamous cell carcinoma reveals novel epigenetically regulated genes and pathways.

Awards

- Skin Care in Organ Transplant Patients Europe (SCOPE): Travel grant to SCOPE meeting in Oslo, Norway, 2010

1. General Introduction

1.1. Cancer biology

"Utque malum late solet immedicabile cancer

Serpere, et illaesas vitiatias addere partes:

Sic letalis hiems paulatim in pectora venit,

Vitalesque vias et respiramina clausit."

“As when a cancer in the body feeds,

And gradual death from limb to limb proceeds;

So does the chilness to each vital part

Spread by degrees, and creeps into her heart;”

Ovid, Metamorphoses

1.1.1. Cancer: Historical Connotations

The term “cancer” has rather famously originated in ancient Greece, where the resemblance of malignant tumours to the shape of a crab (“karkinos”) lead Hippocrates to name malignancies after a sea decapod crustacean (Koss and Melamed, 2006) (Figure 1.1, 1.2). Further latinisation of the word by Celsus of ancient Rome gave grounds to the modern English meaning of cancer: a malignant neoplasm characterised by uncontrolled proliferation of body cells and capacity to invade the host organism.

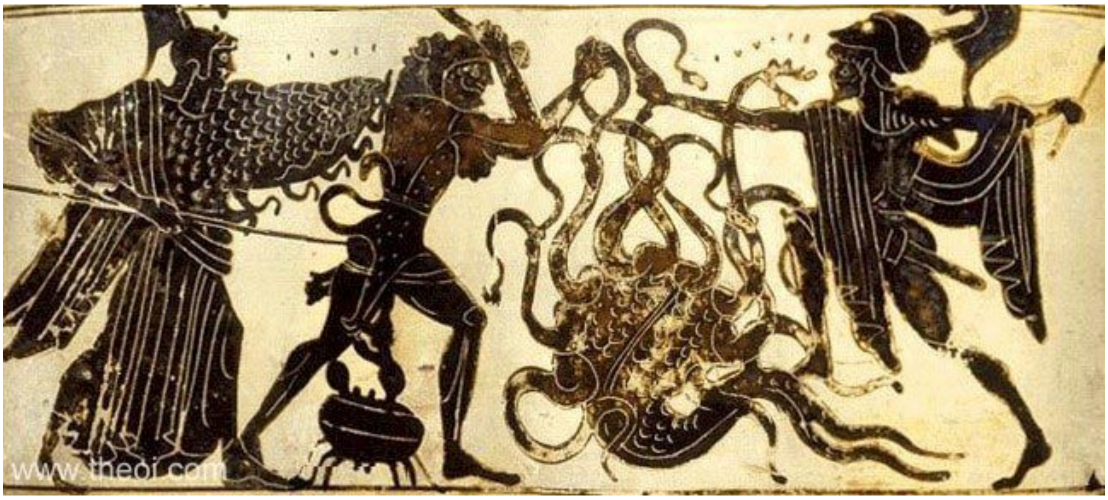


Figure 1.1 Herakles and Hydra. Attic Black Figure by Diosophos (approx. 500 BC). Musée du Louvre, Paris, France, Figure taken from http://www.theoi.com

Figure 1.2 Invasive ductal carcinoma of the breast (macroscopic view). Figure taken from Rosai: Surgical Pathology, 9th edition.

In spite of vast medical progress and therapeutic advances achieved in the past six decades, the term ‘cancer’ in medical context is still, perhaps rightfully, associated with anxiety and fear, and represents a common health concern (Trumbo et al., 2007). Given the recent cancer epidemiology statistics, which suggest that 1.4 million people are diagnosed with cancer each year in the United States alone (Jemal et al., 2006), and one in four deaths are attributable to cancer in the United Kingdom (Statistics, 2010), it may be deduced that cancer is a disease that has reached epidemic proportions, and controlling the onset of novel cases and disease progression in currently affected people is of utmost importance not only from the public health perspective, but also for individual’s welfare: Cancer is associated with considerable mortality

(Ferlay et al., 2007), and also with significant suffering, namely during the advanced stage of the disease (Nilmanat et al., 2010).

Cancer is not a modern disease, although it is often ascribed to modern technology and pollutants it produces, or modern lifestyle (McKinnell, 1998). Conversely, paleopathological evidence suggests that cancer existed already in dinosaurs of the Mesozoic era (250 million years ago) (Rehemtulla, 2010), and has plagued mankind since the dawn of civilisation (David and Zimmerman, 2010). Traits of multiple myeloma (bone marrow cancer) in fossil bones have been discovered in prehistoric American Indians (Morse et al., 1974), nasopharyngeal carcinoma has been reported in an Egyptian mummy (Strouhal, 1978), osteosarcoma has been identified in remains of a Celtic warrior from Bern region in Switzerland (McKinnell, 1998, Ortner et al., 2010), and ample evidence of other malignant tumours has been discovered in skeletons and mummies from many regions of practically all eras (Capasso, 2005).

The value of available historical evidence of cancer epidemiology in the past, albeit limited, is relevant to questions about current cancer rates: Are those on an unequivocal rise due to environmental changes, better diagnostics or rather due to an increase in life expectancy? A study comparing age and gender distribution of cancer between ancient Egypt, modern age (1400-1800) German population and recent English population (early 20th century) suggests that there are no significant differences in cancer distribution when adjusted for age and gender (Nerlich et al., 2006), indicating that prolonged life expectancy is possibly one of the key factors in current trends of cancer incidence. Should current economic growth of Western societies continue, with the accessibility of modern medicine, research and nutrition for the general population, it may be reasonable to presume that life expectancy will continue to rise alongside cancer incidence. This vision underscores the need for effective cancer control and prevention, and unweaving the molecular background and regulation of cancer may provide effective tools for achieving this goal.

1.1.2. Molecular Mechanisms of Cancer

German biologist Theodore Boveri has hypothesised already in 1914 that cancer may be a disease of a single cell originating at the genetic level (Ried, 2009). The genetic element of cancer has been noticed in the Middle Ages, when references to ‘cancer houses’ or ‘cancer families’ were made (Tannock, 2005), but only the recent advances in molecular technologies have been able to clarify the anchoring of cancer in the genome.

Cancer cells acquire intrinsic properties distinct from those of normal cells in the original tissue, and their malignant characteristics are manifested in cellular morphology, inherent functional properties and regulatory mechanisms. The transformation of a normal cell into a malignant one involves complex and dynamic changes of the genome which do not occur simultaneously, but rather represent the outcome of a multi-step process (Hanahan and Weinberg, 2000).

Over two hundred distinct types of cancer have been described and cathegorised, with diverse epidemiology, prognosis, sequelae, risk factors and therapeutic options. In order to define fundamental biological properties shared by various cancers, the following eight cellular features were proposed as key physiological alterations determining malignant growth in the vast majority of tumours: self-sustenance in proliferative signals, unresponsiveness to growth-inhibitory signals, resistance to apoptosis, unlimited potential for replication, sustained angiogenesis, capacity to invade the host tissue, propensity to form distant metastasis, energy metabolism reprogramming, and escaping immune destruction (Hanahan and Weinberg, 2011). Additionally, two so-called "enabling characteristics" were proposed as traits that make the acquisition of the remaining eight features of malignancy possible: genomic instability, which can orchestrate cancerous capabilities by random mutations and chromosomal rearrangements, and tumour-promoting inflammation, capable of promoting tumour progression by various biological means, including the production of growth factors, proangiogenic signals, extracellular matrix-modifying molecules, and by inducing epithelial-mesenchymal transition (EMT).

While the concept of above-described "cancer hallmarks" has been widely accepted as the basis of tumourigenesis, it is important to stress that many of those features, such as sustained angiogenesis or limitless replicative potential, are shared by both malignant and benign tumours (Lazebnik, 2010). While the chief characteristic of cancer cells remains their ability to destroy the host through invasion and metastasis, benign tumours may also occasionally be fatal, and may even share certain molecular characteristics with their malignant counterparts (for example, germline mutation of the RB gene has been described in both retinoma, a benign ocular tumour, and histologically distinct malignant retinoblastoma (Abouzeid et al., 2009)). Additionally, many human malignancies are very lethal while not generating metastases, such as the most common primary brain tumour glioblastoma multiforme. Thus, while those eight physiological traits described as key for malignant transformation may indeed represent crucial determining changes in many if not most cancers, the individual nature of each tumour type (and perhaps tumour subtype) strongly implicates that the relative importance of particular tumourigenic features varies significantly between different malignancies.

From a clinically practical point of view, cancer is essentially a disease of disrupted cell growth and proliferation (Jones and Thompson, 2009). Although cellular proliferation alone does not necessarily lead to cancer, if it is sustained in an environment rich in growth factor, pro-inflammatory factors and agents capable of DNA damage, the neoplastic risk is severely potentiated (Coussens and Werb, 2002). Cells capable of uncontrolled proliferation usually harbour changes in genes and proteins that govern diverse metazoan growth-regulating processes, namely the cell cycle. Growth-regulating processes are evolutionarily conserved, and the molecular heterogeneity of distinct cancers is the mark of stochastic evolutionary forces that vary between individual somatic environments (Evan and Vousden, 2001).

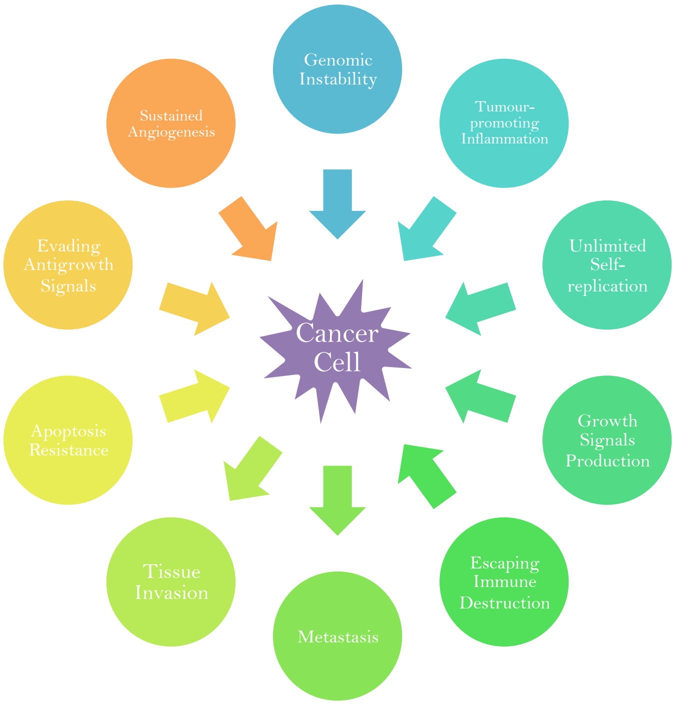


Figure 1.3 Key physiological processes representing currently accepted hallmarks of cancer.

1.1.2.1. Cell Cycle, Tumour Suppressor Genes and Oncogenes

Precise balance between cell division and cell death must be maintained in tissue homeostasis. Mammalian cell cycle represents a process consisting of a series of sequential, highly coordinated steps that involve cell growth, nutrients accumulation, DNA replication and ultimate cell division. The vital importance of this mechanism warrants tight regulation and control mechanisms. Clonal autonomy in vertebrates is restricted by several tiers of regulatory mechanisms, and the disruption of each one must occur for cancers to arise (Evan and Vousden, 2001). Uncontrolled replication alone may not be sufficient for the genesis of human malignancies, and tumour cell resistance to cell death (apoptosis) is another important factor in the tumourigenic process. Unconstrained cell division is nonetheless one of key features of both benign and malignant tumours, thus the dysregulation of the cell cycle is a prominent characteristic of most cancers (Weber and Ashkar, 2000).

Cell cycle is a mechanism regulated by the needs of the whole organism, promoted by cell-extrinsic growth factors and cell-intrinsic protein kinases (Weber, 2007), and inhibited by products of genes that prevent progression of the cell cycle. Those genes belong among tumour-suppressor genes, and include genes such as p21, p27 and p57 (cip/kip family of genes, regulated by a tumour suppressor p53), Rb and p16INK4a and p14arf (Table 1.1). Dysregulation of the cell cycle may occur due to mutations in these cell cycle inhibitors, leading to uncontrolled proliferation and ultimately tumour formation. For example, patients with Li-Fraumeni syndrome, a clinical condition in which p53 germline mutations occur, develop cancer at an early age, and mutations in Rb, the first tumour suppressor gene identified in 1980's, are the molecular cause of retinoblastoma, malignant tumour of the eye retina most commonly diagnosed in young children[1].

Tumour suppressors are not exclusively cell cycle inhibitors, but represent a distinct category of genes. Their translational products include proteins that generally have repressive effects on cell properties that, if uninhibited, would lead to a tumour formation, and are abrogated in cancer cells. Conversely, oncogenes contribute to tumour formation by attributing cellular features that will potentiate the cell to become malignant. The loss of both alleles (two-hit hypothesis) is usually required for the phenotype of tumour suppressor loss (Rook and Burns, 2010), while a single allele loss of an oncogene can be sufficient for phenotypic presentation, although there are exceptions to this rule.

Oncogenes are evolutionarily conserved molecules, mostly belonging among growth factors, growth factor receptors (GFR) or down-stream effectors of GFR (Neal and Hoskin, 2009). It is generally estimated that between five to seven genes must be abolished in a cell to develop cancerous phenotype, and those must include both activated oncogenes and inactivated tumour suppressors, yet there are tumours in which a single genetic mutation represents the direct cause of the disease: the vast majority of chronic myeloid leukemia cases are characterised by a cytogenetic abnormality t(9, 22) translocation, the so-called "Philadelphia chromosome". This genetic mutation leads to the formation of the BCR-ABL oncogene, resulting in subsequent dysregulation of molecular signalling which enhances proliferative potential and apoptosis resistance in progenitor cells of the haematopoietic system. The ultimate result is a dramatic increase in the number of myeloid cell (Arora and Scholar, 2005).

Both oncogenes and tumour suppressors are regulated by genetic and epigenetic mechanisms. Genetic changes include changes in the DNA sequence by mutation (base substitution, insertion, or deletion, chromosomal rearrangement), or aberrant copy number of DNA segments (Stratton, 2011), while epigenetic changes occur mainly due to three mechanisms: promoter methylation or demethylation, histone modification and microRNA (miRNA) dysregulation. In summary, the genetic cause of cancer can be a single driving change (such as mutation), or more likely a combination of genetic and epigenetic changes capable of causing cellular phenotype that will lead to a tumour formation.

1.1.2.2. Retinoblastoma gene

Since its seminal discovery as the first tumour suppressor gene in 1986 (Friend et al., 1986, Lee et al., 1987), the mechanistic characterisation of the RB gene has focused namely on its function in cell cycle progression control. Its capacity to arrest cell cycle in the G1 phase was thought to represent the main mechanism through which RB executes its tumour suppressor function. However, it has become clear that in addition to G1 checkpoint, RB posseses crucial capacity in

---

[1] As described above, this mutation is detected also in retinoma, a benign tumour of the eye.

other processes, including tissue differentiation during both embryogenesis and adult life, apoptosis, cell cycle arrest maintenance and preserving chromosomal stability (Burkhart and Sage, 2008).

Rb protein (pRb) is able to bind and thus inhibit E2F transcription factors during the G1 phase. E2F regulates several downstream genes, many of which are involved in key cell cycle regulatory processes. The expression of E2F target genes is repressed by this complex through recruiting additional corepressors that modulate chromatin (Gordon and Du, 2011). Those include histone deactylases, DNA methyltransferase 1 (DNMT1), and heterochromatin protein 1A (HP1A) (Burkhart and Sage, 2008).

The coupling of pRb and E2F is inhibited by various cyclin-dependent kinases (CDKs). Mitogens produced during G1 phase induce activation of cyclin D/Cdk4-Cdk6, and cyclin E/Cdk2, which leads to pRb phosphorylation and E2F release. Upon activation of downstream genes, the cell progresses from G1 to S phase. The phosphorylation of pRb and subsequent cell cycle progression is conversely inhibited by antigrowth signals through down-regulating cyclin protein levels, or by inducing cyclin-dependent kinase inhibitors, such as INK4A or CIP/KIP family genes. Thus, the Rb protein serves as a cell cycle “switch” that is controlled by both growth promoting and growth inhibiting molecules.

Dysregulation of this pathway has been described in many cancers, including sporadic breast, bladder, prostate, lung, liver, oesophageal, brain cancer, melanoma and leukaemia. The mechanisms of RB inactivation include not only DNA mutation, but also hyperphosphorylation by overexpressed cyclins and CDKs, DNA methylation, miRNAs, and sequestration by viral oncoproteins. In cervical squamous cell carcinoma (SCC) and head and neck SCC, human papillomavirus (HPV) 16 is suspected to induce malignant transformation by inactivating RB via the expression of the E7 oncoprotein (Wittekindt et al., 2012).

The currently known communication network of pRb is vast and still expanding, and contains many binding partners and target proteins involved in diverse cellular proceses (Figure 1.4). This clearly demonstrates that the role of the RB gene is by no means linear, and its precise contribution to cancer initiation, progression and treatment resistance likely varies from tumour to tumour.

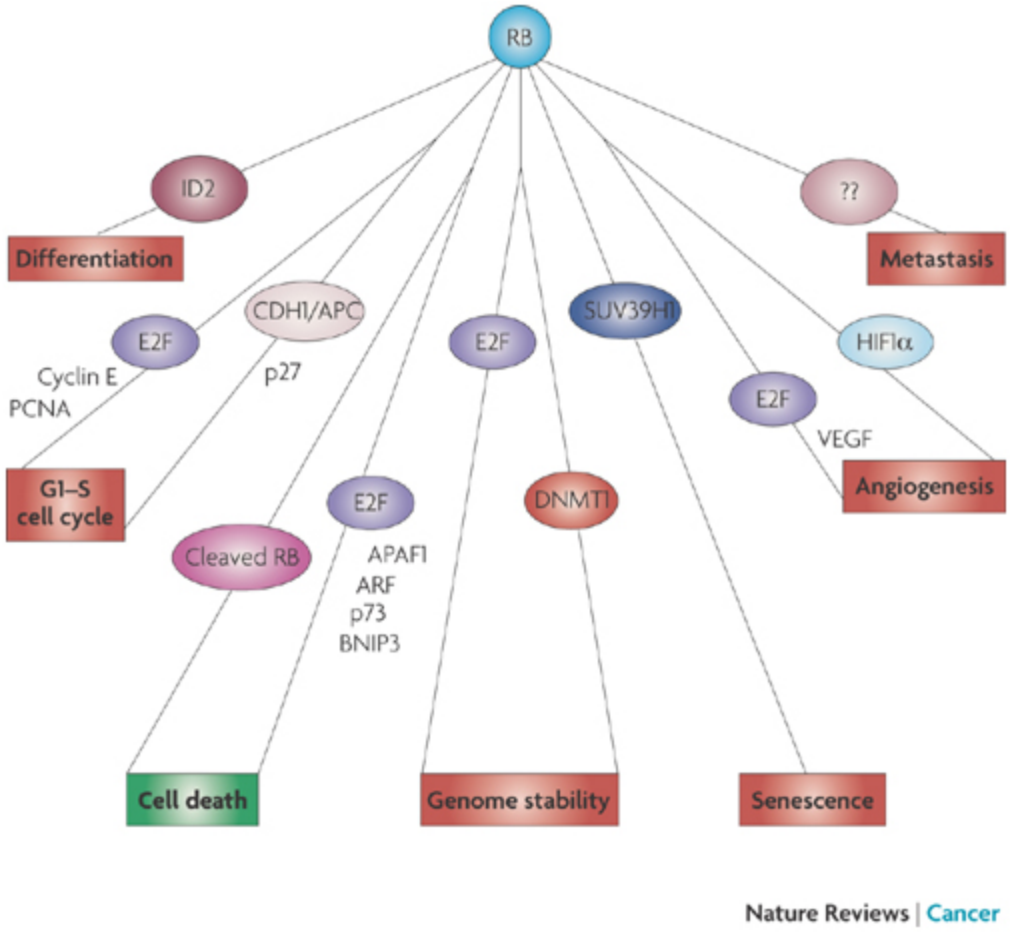


Figure 1.4 The role of RB in various critical cellular processes. Figure taken from (Burkhart and Sage, 2008).

1.1.2.3. P53 gene family

The P53 gene, also known as the “guardian of the genome”, is the most commonly mutated gene in human cancers. One of its key functions as an archetypal checkpoint regulator is to ensure that a replicating cell arrests in G1 phase in response to DNA damage, and repairs its genome prior to cell cycle progression (Sherr, 1996). The loss of p53 leads to decreased fidelity during DNA replication, aberrant chromosome segregation due to multiple centrosomes, and unless changes are severe enough to induce a mitotic catastrophe, the ultimate result is a rapidly proliferating cell with a highly unstable genome, resistant to apoptosis and capable of progession towards malignancy.

In addition to regulating cell cycle progression, the plethora of p53 downstream effectors regulate other important mechanisms such as cellular metabolism, senescence, cell motility, and to some effect response to cancer therapy (Lai et al., 2012, Barbieri et al., 2006). Generally speaking, cells with intact p53 show greater response to anticancer therapies (O'Connor et al., 1997).

Similar to the RB gene, inactivation of P53 occurs through multiple mechanisms distinct in various tumour types. Mutations in the DNA-binding domain are typical for colon, breast, lung, bladder, brain, pancreas and other cancers, inactivation by viral oncogenes has been described in cervical cancer and lymphomas, p53 mislocalisation to the cytoplasm leads to breast cancer and neuroblastoma (Vogelstein et al., 2000). Epigenetic modifications of P53 include DNA methylation, targeting by miRNA and by non-coding RNA, and chromatin organisation (Saldana-Meyer and Recillas-Targa, 2011).

P63, described in 1997 as a member of p53 family, is a molecule that posseses both tumour suppressor and oncogenic properties. Its full length isoform contains the transactivation domain (TAp63), and is capable of inducing cell cycle arrest and apoptosis (Gressner et al., 2005). On the other hand, the amino-deleted p63 isoforms (ΔNp63) play an important role in development and epidermal organisation, and exhibit oncogenic properies (Yang et al., 1999, Mills et al., 1999).

In contrast to p53, mutations of p63 are rare in human cancers. In fact, p63 can even be overexpressed in certain cancers, such as bladder cancer or cervical carcinoma. Many tumours show simultaneous upregulation of both TAp63 and ΔNp63 isoforms, with the latter being more abundant at the protein level (Melino, 2011). While the ratio of individual isoforms seems to determine the cell fate, the loss of endogenous p63 leads to increased invasiveness and metastatic potential of squamous cell lines, and the acquisition of mesenchymal properties (Barbieri et al., 2006).

P73 bears many similarities to P63: it is rarely mutated in cancers, and also transcribed in various isoforms which display antagonistic effects; the TAp73 isoform shows properties analogous to tumour suppressor gene by inducing cell cycle arrest, apoptosis and protecting against genomic instability, while the ΔNp73 isoform has oncogenic features (Conforti et al., 2012). The ultimate regulatory effect imposed by both p63 and p73 seems to depend on precise ratio of various isoforms and other transcription factors. Precise details of this process remain to be elucidated, and may serve as the basis of future targeted cancer therapies.

Other important tumour suppressor genes and malignancies associated with their hereditary alterations are listed in Table 1.1.

Gene	Associated hereditary syndrome	Major hereditary tumour types	Chromosomal location
APC	FAP	Colon, thyroid, stomach, intestine	5q21
ATM	Ataxia telangiectasia	Leukaemias, lymphomas, brain cancer	11q22-q23
BHD	Birt-Hogg-Dube syndrome	Kidney, hair follicle tumours	17p11.2
BLM	Bloom syndrome	Leukaemias, lymphomas, skin cancer	15q26.1
BRCA1 and BRCA 2	Hereditary breast cancer	Breast, ovarian cancer	17q21 13q12.3
CDKN2A	Familial malignant melanoma	Melanoma, pancreas	9p21
CYLD	Familial cylindromatosis	Pilotrichomas	16q12.1
FANCA	Fanconi anaemia	Leukaemias	16q24.3
FH	Hereditary leiomyomatosis	Leiomyoma	1q42.1

FHIT	Muir-Torre familial cancer syndrome.	Non-melanoma skin cancer, colon cancer, renal cell carcinoma	3p14.2
GPC3	Simpson-Golabi-Behmel syndrome	Embryonal cancers	Xq26.1
MEN1	Multiple endocrine neoplasia type I	Parathyroid, pituitary cancer, carcinoid	11q13
NF1	Neurofibromatosis type 1	Neurofibroma	17q11.2
NF2	Neurofibromatosis type 2	Meningioma, acoustic neuroma	22q12.2
PTCH	Gorlin syndrome	Skin, medulloblastoma	9q22.3
PTEN	Cowden syndrome	Hamartoma, glioma, uterus	13q14.2
RB1	Hereditary retinoblastoma	Retinoblastoma	13q14
STK11	Peutz-Jeghers syndrome	Intestinal, ovarian, pancreatic cancer	19p13.3
SUFU	Medulloblastoma predisposition	Skin, medulloblastoma	10q24.32
TP53	Li-Fraumeni syndrome	Breast, brain, leukaemia, adrenal cancer	17p13
TSC1, TSC2	Tuberous sclerosis	Hamartoma, kidney	9q34 16p13.3
VHL	Von Hippel–Lindau syndrome	Kidney	3p26-p25
WT1	Familial Wilms Tumour	Kidney tumour	11p13
XPA	Xeroderma pigmentosum	Skin cancer	9q22.3

Table 1.1 Tumour suppressor genes important in human malignancies. Table adapted from (Vogelstein and Kinzler, 2004).

1.1.2.4. Oncogenes

The term oncogene denotes a gene capable of contributing to malignant transformation under certain conditions, such as mutations or viral infection. Proto-oncogene is a gene that becomes tumourigenic if its expression becomes excessive, or if it is consitutively activated, as opposite to its wild-type counterpart. Oncogene activations can occur due to chromosomal translocations (such as the above-described Philadelphia chromosome in CML), gene amplifications, or due to intragenic mutations that alter residues with regulatory function of the translated protein (such as BRAF mutation changing a valine to a glutamate). In contrast to tumour suppressor genes, an activating somatic mutation in a single oncogene allele suffices to provide a selective growth advantage to the cell (Vogelstein and Kinzler, 2004). Many oncogenes were first identified during studies of virus-induced tumourigenesis (see Chapter 1.1.4.3), and retrovirus-associated oncogenes relevant in human tumours are listed in Table 1.2.

Gene	**Mechanism of inactivation**	**Associated tumour types**	**Chromosomal location**
MYC	Amplification	Neuroblastoma, medulloblastoma	8q24.21
	Increased transcription	Burkitt's lymphoma, B-cell lymphoma	
EGFR	Mutation	Glioblastoma, non-small-cell lung cancer	7p12
RAS	Mutation	Pancreatic cancer	12p12.1 (KRAS)

RAF	Mutation	Melanoma	7q34 (BRAF)

Table 1.2 Oncogenes first identified in retroviruses as drivers in human cancer. Table adapted from (Vogt, 2012).

1.1.2.5. MYC Oncogene

The MYC oncogene is a DNA-binding protein known to contribute to the formation of many malignant tumours. The gene was first shown to be altered by chromosomal translocation in Burkitt lymphoma (Dalla-Favera et al., 1982), and since then found to be translocated or amplified in an array of cancers, including multiple myeloma, T-cell leukemia and colon carcinoma.

MYC interacts with many canonical pathways, including TGFβ, Wnt, Hedgehog, Notch, p53, PI3K, and its dysregulation thus disrupts many cellular processes, including metabolism and cell cycle. Importantly, MYC is capable of inducing many miRNAs, including mir-17 cluster which inbihits tumour suppressors such as PTEN (Dang, 2012).

In addition to its implication in the genesis of cancer, MYC was described as one of four genes that are capable of reprogramming differentiated fibroblasts to a pluripotent stem cell (Takahashi and Yamanaka, 2006). The precise molecular role of MYC in pluripotent reprogramming is not entirely clear, but it seems to involve chromatin structure regulation and polycomb-mediated gene silencing. However, it is of note that in contrast to its function in pluripotency maintenance, MYC is required for terminal differentiation of human epidermal stem cells (Gandarillas and Watt, 1997). This data suggests that the concrete role of MYC depends on given cellular context and the balance of other regulatory molecules.

Other important oncogenes are listed in Table 1.3.

Gene	Associated hereditary syndrome	Major hereditary tumour types	Chromosomal location
KIT	Familial gastrointestinal stromal tumours	Gastrointestinal stromal tumours	4q11-q12
MET	Hereditary papillary renal cancer	Kidney cancer	7q31
PDGFRA	Familial gastrointestinal stromal tumours	Gastrointestinal stromal tumours	4q12
RET	Multiple Endocrine Neoplasia Type 2	Thyroid, parathyroid, adrenal cancer	10q11.2

Table 1.3 Important oncogenes frequently mutated in human malignancies. Table adapted from (Vogelstein and Kinzler, 2004).

It seems probable that the inhibition of overexpressed oncogenes and functional restoration of mutated tumour suppressor genes may abrogate malignant properties of cancerous cells. Many compounds that target specific molecular alterations in cancer are currently approved for application in clinical practice (see Chapter 1.1.3), and many more are currently being scrutinised in preclinical testing and clinical trials. However, the complexness of cellular signalling network suggests that the inhibition of a single molecule is unlikely to achieve full anticancerous effect, and that multimodal approach to treatment represents the best clinical practice.

1.1.3. Cancer therapies based on molecular properties of target tumours

The fundamental goal of cancer therapy is to permanently eradicate malignant cells, while preserving normal cells intact. The non-selective nature of traditionally administered therapeutic agents including cytotoxic chemotherapy and radiation treatment, while unequivocally therapeutically beneficial, often leads to significant damage in healthy tissues, namely in those which contain rapidly dividing cells, as this cellular feature is directly targeted by those therapeutic modalities. The most common tissues adversely affected by chemotherapy therapy include the skin, cutaneous adnexa, mucous membranes, gastrointestinal tissue and the haematopoietic system. Additionally, radiation treatment may cause damage to healthy tissues its beam has to penetrate in order to reach the target tumour mass.

An important trait of many tumours is their resistance to various anticancer agents; either intrinsic or acquired during treatment course. The most frequent causes of acquired resistance to a broad range of anticancer chemotherapeutics (multidrug resistance) include the expression of energy-dependent transporters that recognize and subsequently eliminate chemotherapeutics from the cell, resistance to drug-induced apoptosis, and the initiation of drug-detoxifying mechanisms (Gottesman, 2002).

Unravelling molecular mechanisms driving tumourigenesis holds a great promise for personalised cancer therapy; however, practically all patients become resistant to those agents at a certain stage of their disease. Specific cell-autonomous molecular mechanisms, such as MET overexpression, and tumour microenvironment alterations confer resistance to molecular targeted therapies (Straussman et al., 2012), suggesting strong compensatory mechanisms are induced within cancer cells and the surrounding space in order to overcome the anticancerous effect of those drugs.

Many cancer therapies are designed to target a specific molecular mechanism driving cancer cells: radiation therapy targets mainly rapidly dividing cells due to cell cycle dysregulation, azacitidine used in the treatment of myelodysplastic syndrome (Kaminskas et al., 2005) reduces DNA methylation (Antonsson et al., 1987), targeted therapeutic agents include small molecules that inhibit a specific protein that promotes cancer (Mladkova and Chakravarti, 2009). Examples include imatinib, which inhibits bcr-abl protein in chronic myeloid leukaemia (Deininger et al., 2005), bevacizumab, which inhibits vascular endothelial growth factor (Mladkova and Chakravarti, 2009) and is approved for application in colon, breast, brain, kidney and lung cancer (Cohen et al., 2007, Pazdur, 2010), rendering obvious the fact that discovering molecules responsible for cancer onset and growth is a feasible route towards novel therapies and potentially better perspectives for cancer patients.

A list of selected targeted therapeutics currently approved for clinical application is listed in Table 1.4.

Drug name	Molecular target	Approved indication
Bevacizumab	VEGF	Metastatic kidney and colon cancer, glioblastoma, lung cancer
Brentuximab vedotin	CD30	systemic anaplastic large cell lymphoma, Hodgkin lymphoma
Cabozantinib	VEGF receptors, RET, MET, TRKB, and TIE2.	Metastatic medullary thyroid cancer
Cetuximab	EGRF, HER-1	Squamous cell carcinoma of the head and neck, colorectal cancer
Erlotinib	EGFR	Non-small cell lung cancer and pancreatic cancer
Ipilimumab	CTLA-4	Metastatic melanoma
Pazopanib	VEGF receptors, c-KIT, PDGFR	Advanced renal cell carcinoma and advanced soft tissue sarcoma.
Trastuzumab	ERBB2	ERBB2-positive breast cancer, ERBB2-positive gastric or gastro-oesophageal junction carcinoma
Vemurafenib	BRAF V600E	Metastatic melanoma
Vismodegib	SMO	Metastatic or recurrent locally advanced basal cell carcinoma
Vorinostat	HDAC inhibitor	Cutaneous T-cell lymphoma

Table 1.4 Targeted therapeutics currently approved for clinical application in cancer patients.

1.1.4. Pathogenesis of cancer

The mechanism by which various oncogenes or tumour-suppressors are dysregulated varies considerably between tumour to tumour. However, epidemiologal data provide evidence for some general mechanisms responsible for tumour formation: those may result due to the action of an environmental factor or due to a leverage of microorganisms.

1.1.4.1. Chemical carcinogens

Chemical factors epidemiologically linked to cancer include industrial products, but also the products of microorganisms. Examples include cigarette smoking, responsible for approximately 25% of cancer-related death in the UK (Neal and Hoskin, 2009), leading mainly to lung, oral cavity and larynx cancer. Exposure to asbestos is linked to mesothelioma (Peto et al., 1995), aniline dyes are associated with cancer of the urinary system (Clayson, 1981), and soot captured in the skin rugosity of the scrotum of chimney sweeps leads to cutaneous squamous cell carcinoma (Johnson et al., 1992). Aflatoxin produced by Aspergillus flavus, a fungal contaminant of poorly stored food, causes hepatocellular carcinoma (Liu and Wu, 2010), and nitrosamines, which are the product of intestinal microflora from nitrogenous compounds, have been implicated in gastric cancer (Hill et al., 1973).

1.1.4.2. Physical carcinogens

Physical factors are capable of causing DNA damage, such as strand breaks or point mutations. Such events occur in human organism on a daily basis; however, the damaged DNA is either repaired or the cell is removed and prevented from multiplication. If those mechanisms fail, the damaged cell can replicate and promote the genetic error, possibly leading to cancer formation. Given that skin and mucosal surfaces are most exposed to the external environment, tumours resulting from physical carcinogens are especially common in those areas.

Ultraviolet radiation: Solar radiation is created in the core of the Sun by nuclear fusion, when two hydrogen atoms are fused into helium. This process causes atoms to discharge photons, which are absorbed by the adjacent gas molecule (about 1 mm away), and upon photon absorption, the gas molecule heats up and releases a photon, which then undergoes the same process until it reaches the Sun surface. The time it takes the photon to reach the surface via repeated absorptions and re-emissions is about 100,000 years, and another 8 minutes to reach the Earth. Approximately 40% of solar radiation that reaches the atmosphere passes down to the earth's surface. In the atmosphere, ultraviolet (UV) radiation < 200 nm in wavelength (vacuum UV) is absorbed by oxygen and nitrogen and turned into heat, UV radiation with wavelengths from 200 to 300 nm (UVC, most of UVB) is absorbed by ozone (O3), infrared radiation with wavelength > 700 nm is partially absorbed by carbon dioxide, ozone, and water vapours. About 30% of the sun's visible spectrum (wavelengths from 400 nm to 700 nm) is reflected back to space. Ultimately, 98% of UV radiation that reaches the Earth's surface is UVA radiation.

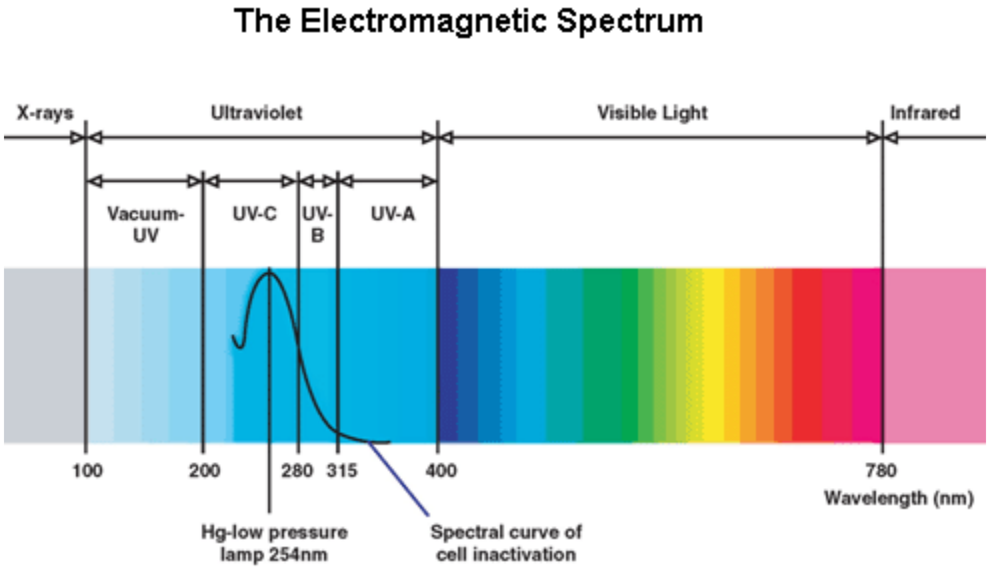


Figure 1.5 The electromagenetic spectrum. Figure taken from http://www.eclogiteskincare.com/blog/wp-content/uploads/2009/07/uv-light-spectrum.gif.

The most destructive from of UV radiation is the UVB, albeit needed for vitamin D synthesis, since it causes photochemical damage to DNA. Excessive exposure to UVB radiation can lead to actinic keratoses (precursors of cSCC – see section 1.3.2.2), cSCC, BCC and melanoma.

Sunbeds used for cosmetic tans emit about 3% of UVB and have been linked to an increased risk of skin cancer, and have been classified as a human carcinogen by by the International Agency for Research on Cancer. Moreover, the vast majority of sunbeds in England exceed safe levels of UV radiaton, increasing the risk of skin cancer for users (Tierney et al., 2013).

Ionising radiation consists of particles capable of detaching electrons from atoms ('ionising' them). There are both natural and artificial sources of radiation, including cosmic radiation or Radon gas, and nuclear power-plants or nuclear medicine. The main effect of ionising radiation includes DNA damage, particularly in dividing cells, which can lead to tumour formation. Hiroshima and Nagasaki atomic bomb survivors developed leukaemia and many solid cancers (Ozasa et al., 2011), Chernobyl accident lead to thyroid cancer namely among children in the contaminated areas (Reiners et al., 2008), and political prisoners in the former Czechoslovakia sentenced to uraninite mining mostly developed lung cancer or leukaemia (Kulich et al., 2011).

1.1.4.3. Oncogenic viruses

Viruses are non-cellular organisms unable to reproduce without the host cell. Given that a considerable proportion of cancers - estimated at 15% (Parsonnet, 1999) - is linked directly or indirectly to viral infection, future preventive measures including vaccines may significantly contribute to decreasing the incidence of at least some cancers.

Retroviruses

Retroviruses, a group of RNA viruses characterised by reverse transcription, played a seminal role in the elucidation of many biological processes, including the molecular mechanism of malignant transformation. Following infection of the host cell, the viral RNA genome is transcribed by the enzyme reverse transcriptase into a double stranded DNA. This genomic product is then integrated into the host genome at almost random sites with the help of the enzyme integrase, becoming a provirus (Butel, 2000). The infection is permanent, as the provirus is practically never lost from the genome. The provirus may remain inactive, or be actively transcribed by the cellular enzyme RNA polymerase II, which ultimately leads to the production of new viruses that can infect other host cells, or other hosts. If the virus infects a germline cell, it becomes an "endogenous retrovirus" and can be inherited by subsequent generations through classical Mendelian inheritance (Groom and Bishop, 2012)[2].

Historically, the very first retrovirus isolated from a neoplastic disease came from the blood of chickens affected with erythro-myeloblastic leukaemia, and the filtered product was reported to induce the disease in inoculated chickens (Rubin, 2011a). In 1911 Peyton Rous isolated "a separable agent" from the spindle cell sarcoma of a Plymouth Rock hen, later called Rous sarcoma virus (RSV). This agent was also capable of inducing solid tumours in other inoculated

---

[2] Some 8 % of the human genome are believed to consist of retroviral sequences.

birds (Rous, 1911), and served as the stepping stone for decades of cancer research to follow. In subsequent years, tens of retroviruses were isolated from various animal tumours (mainly avian and murine), and eventually it was discovered that malignant transformation incited by retroviruses is induced through oncogenes; either endogenous to the virus, or by the acquisition of protooncogenes of the cell. Based on this determining feature, oncogenic retroviruses fall into two groups: acute transforming retroviruses, and slow transforming retroviruses.

Acute transforming retroviruses, as their name implies, transform the host cell rapidly because of carrying additional genetic information – viral oncogenes (Duesberg and Vogt, 1970). Their genome is simple, incapable of viral self-replication, and their high tranformation capacity leads to tumours within weeks of infection. The RSV is a prototype of this class, and its *src* oncogene encodes a tyrosine protein kinase critical for regulating signal transduction. Approximately 25 viral oncogenes were discovered to date, and it is of important note that these retroviral oncogenes were originially derived from normal cellular genes (Butel, 2000). Essentially, these retroviruses contain a viral homologue of a cellular proto-oncogene that was “stolen” from the host cell during phylogenesis of the retrovirus to promote its replication.

In contrast to acute transforming retroviruses, slow transforming retroviruses have more complex genomes, are not replication-defective, do not carry their own oncogenes and tumours they induce require longer periods of time to form. Malignant transformation occurs by several mechanisms: cellular protooncogenes are activated by the insertion of a viral long terminal repeat (LTR) close to a protooncogene promoter or enhancer, causing overexpression of the cellular gene, which then becomes responsible for the tumour formation (Fan and Johnson, 2011). More than 70 proto-oncogenes have been described to be activated by the insertion of a retrovirus. Other oncogenic mechanisms include coding for auxillary viral oncoproteins (such as *tax* protein coded by HTLV-1), and retrovial envelope (env) proteins that cause lung cancer in sheep infected by JSRV.

In humans, important cancerogenic retroviruses include Human T-cell lymphotropic virus type 1 (HTLV-1) and HTLV-2. HTLV-1 causes adult T-cell leukaemia-lymphoma in Japan and Caribbean regions, HTLV-2 infection can lead to haematologic malignancies. While not directly oncogenic, HIV-1 and HIV-2 are retroviruses that cause AIDS in humans, a disease associated with increased cancer risk, and thereby indirectly contribute to cancer indicence world-wide.

DNA viruses

Genomes of small DNA oncogenic viruses (polyomaviruses, papillomaviruses, adenoviruses) do not posses physiological properties essential for the genome replication, nor the components necessary for genomic transcription. Thus, malignant transformation by DNA viruses requires multiple steps that lead to the transition of a quiescent cell into a proliferating one, and subsequent utilisation of the host cell’s transcription and replication machinery by the virus. The proliferative state is induced by the products of transforming oncogenes, which are of viral origin in contrast to retroviral oncogenes, and are indispensible for viral replication and other functions.

DNA viruses hijack the cell cycle for its own replication by targeting key regulatory molecules including tumour suppressors pRb and p53. Viruses achieve this goal through various means: upon binding to p53, SV40 large T antigen (T-ag) inhibits the DNA binding activity of p53, while HPV oncoprotein E6 leads to ubiquitin-mediated degradation of the target protein (Butel, 2000). Both T-ag and HPV E7 contain a short region of homology capable of binding to pRb, with selectivity towards its active (hypophosphorylated) form. It is of note that DNA oncogenic viruses share the ability to inhibit pRb and p53 although they are not evolutionarily too closely related. An oncogenic DNA virus may cause a single type of cancer, or multiple tumour types. Moreover, some viruses are able to cause tumours in a variety of tissues, and even non-neoplastic diseases (see Table 1.5). Given that not all infected hosts develop tumours, it seems very likely that many host factors including the immune state play an important role in virus-induced tumourigenesis.

Other oncogenic viruses include Hepatitis B and C viruses which cause hepatocelular cancer; Epstein-Barr virus (human herpes virus 4), which causes infectious mononucleosis, has been linked with lymphoma formation, namely in Africa (Parkin, 2006); human herpes virus 8 is the causative agent of Kaposi’s sarcoma (Moore and Chang, 1995).

Recently, a novel DNA polyoma virus has been discovered in samples of Merkel Cell carcinoma (MCC), a rare type of skin cancer, and is suspected to be the cause of majority of MCC. It is the first example of polyoma virus integrating into the host genome (discussed in more detail in Chapter 5).

Summary of known oncogenic viruses is provided in Table 1.5.

Characteristic	**HBV**	**EBV**	**HPV**	**HHV-8**	**SV40**	**MCPyV**
Genome size (kilobasepairs)	3.2	172	8	165	5.2	5.4
Cell tropism	Hepatocytes, white blood cells	Oropharyngeal epithelial cells, B cells	Squamous epithelial cells (mucosal, cutaneous)	Vascular endothelial cells, lymphocytes	Kidney epithelium, other tissues	Unclear
Transmission	Vertical, parenteral, horizontal, venereal	Saliva	Venereal, skin abrasions	Horizontal, venereal	Faecal-oral	Unclear
Human diseases	Hepatitis, cirrhosis	infectious mononucleosis, oral hairy leukoplakia	Skin warts, EV, genital warts, laryngeal papillomas			

Associated tumours	Liver cancer	nasopharyngeal carcinoma, lymphomas	Cervical, skin, oropharynx	Kaposi's sarcoma, PEL,Castleman's disease	Brain, bone, mesothelioma	MCC, non small cell lung cancer?

Table 1.5 Key ds-DNA oncogenic viruses and their characteristics. PEL - primary effusion lymphoma. Table adapted from (Butel, 2000).

HPV

Human papillomavirus (HPV) is a double-stranded DNA virus with more than 100 types characterised based on L1 gene homology (figure 1.4). HPV currently represents one of the most common sexually transmitted infections globally, and the majority of sexually active individuals become infected during their lifetime (Baseman and Koutsky, 2005). However, most infected individuals experience an asymptomatic infection course, with the virus clearance occurring in up to 90% of cases, and only the remaining 10% individuals with persistent infection are at increased risk of developing cancer (Zandberg et al., 2013).

Commonly, HPV causes viral warts (Orth et al., 1977) (Figure 1.5), but a subset of α-HPV, including types 16 and 18, confer a high risk of malignant transformation and account for up to 70% of cervical cancers (Parkin, 2006), which in recent years has lead to the development of a clinically approved vaccine against those HPV types (Saslow et al., 2007). In non-melanoma skin cancer, the mechanistic link between HPV infection and tumour formation is still rather elusive (see section 1.3.4).

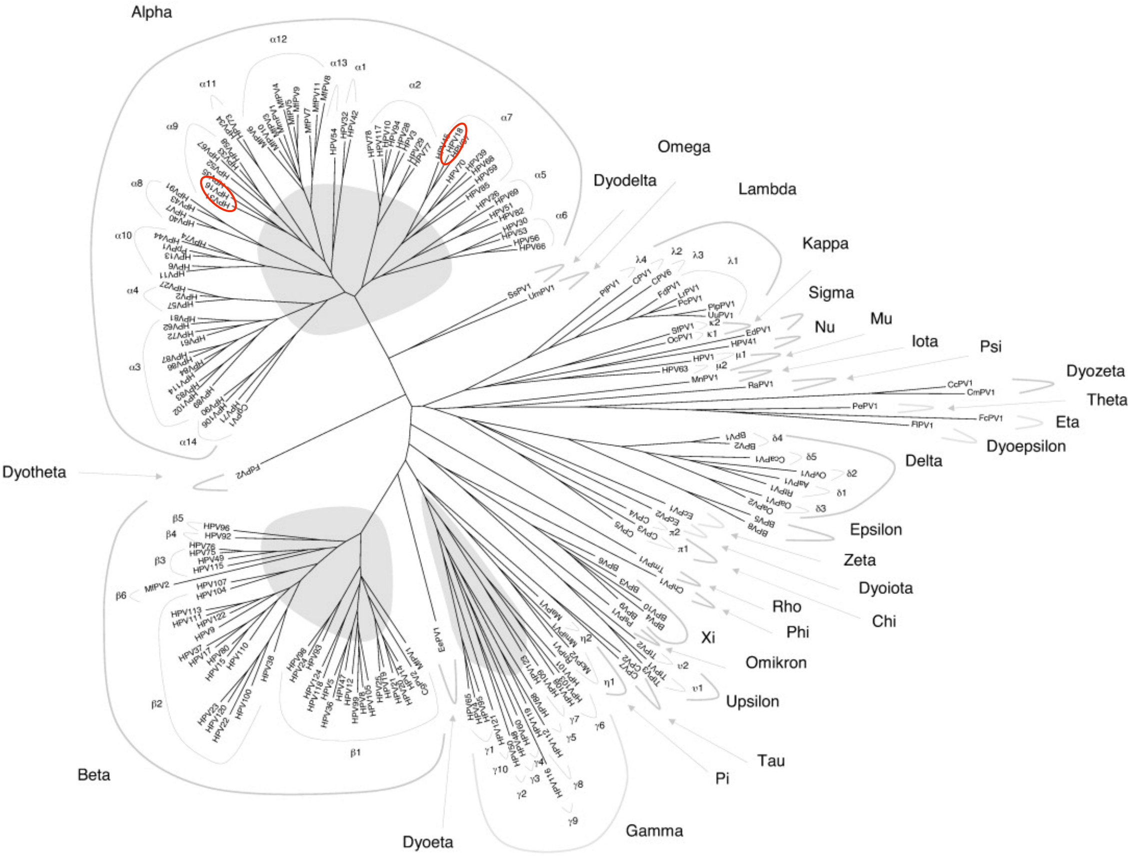


Figure 1.6 Phylogenetic tree of papillomaviruses according to LT1 gene sequence of 189 papillomaviruses. Figure taken from (Bernard et al., 2010).

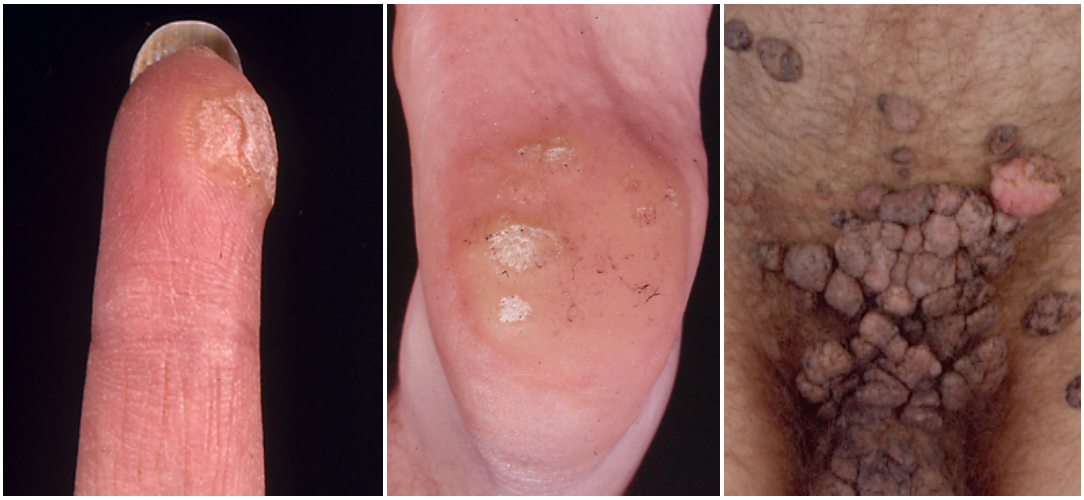


Figure 1.7 Clinical presentation of viral warts in various anatomic locations.

HPV and genital and non-genital cancers

The aetiological link between an infectious agent and genital cancers has been first suspected in 1842, when Rigoni-Stern, an Italian doctor, noticed an epidemiological association between the mortality from cervical cancer and sexual activity. Over a century later, HPV particles were first demonstrated in cervical condylomata in 1978, and in 1980, Gissmann and zur Hausen isolated HPV DNA from genital warts (Gissmann and Hausen, 1980). It was later shown that HPV genotype 16 (HPV-16) and HPV-18 represent key subtypes directly linked to cervical cancer onset. In recent years, the association of HPV and human cancer has broadened from genital tumours to neoplasia of other anatomical locations (see Table 1.6), including anal, vaginal, vulvar and penile cancers, in addition to oropharyngeal cancer (Zandberg et al., 2013). This epidemiological trend suggests that HPV vaccines may be effective in preventing malignancies in diverse groups of patients.

Anatomic area	Annual no. of cases	HPV-associated	HPV 16/18 associated
Cervix	11,845	11,370	9000
Vagina	714	460	400
Vulva	3062	1560	1350
Anus and rectum (women)	2977	2770	2590
Oropharynx (women)	2306	1450	1380
Total (women)	20,903	17,61	14,72
Penis	1000	360	310
Anus and rectum (males)	1618	1500	1410
Oropharynx (males)	8936	5630	5360
Total (males)	11,553	7490	7080

Table 1.6 Tumours attributable to HPV Infection in the United States From 2004 to 2007. Table taken from (Zandberg et al., 2013)

Biology of HPV infection

HPVs are small DNA viruses that lack envelope. Their double-stranded genome is divided into early (E) and late (L) regions, and encodes seven non-structural “early” proteins (E1, E2, E3, E4, E5, E6 and E7) and two “late” structural proteins (L1 and L2), which code for proteins that constitute the icosahedral viral capsid. Additionally, there is a non-coding regulatory long control region (LCR).

E genes modulate cell functions and regulatory circuits in favour of viral reproduction. Specifically, their products alter keratinocyte differentiation, promote apoptosis resistance, induce viral DNA replication and thwart immunological clearance of the virus. It is of note that the HPV genome is present in the host cell in the form of an episome, and its integration into the host genome is rather rare and not necessary for HPV replication (Yuan et al., 2012)

The life cycle of HPV is closely linked with the differentiation of infected epithelial tissue: the virus infects cells in the basal layer of stratified epithelia, and its genome becomes a nuclear episome replicated along with nuclear chromosomes. During this early phase following HPV infection, the E1 and E2 genes are transcribed in order to regulate viral replication. As the infected cell divides, becomes more differentiated and migrates towards suprabasal epithelial layers, L genes are activated, major and minor capsid proteins they produce encapsidate viral genomes and a new generation of virions is produced. The E6 and E7 oncogenes are expressed during this phase, and affect the cell cycle to stimulate differentiating cells towards a proliferative state in order to keep replication of the viral genome going (Longworth and Laimins, 2004). While the oncogenic process is doubtless complex and involves many host factors, dysregulation of the cell cycle, abrogated apoptosis, inhibited cellular differentiation and chromosomal instability represent key HPV-induced features that contribute to the process of malignant transformation.

Various HPV types are described as “low-risk” or “high-risk” from the perspective of their oncogenic potential. So-called high-risk HPV types include previously mentioned HPV-16 and 18, and also 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82. Low-risk types include HPV-6, 11, 40, 42, 43, 44, 53, 54, 61, 72, and 81. These HPV types cause less severe lesions, including genital condylomata and laryngeal papillomas (Steben and Duarte-Franco, 2007).

In most regions, high-risk HPV-16 is more prevalent than other high-risk HPV types, suggesting perhaps higher virulence or cellular tropism of this subtype. Moreover, HIV-positive patients co-infected with HPV demonstrate a considerably increased rate of HPV 16-positive tumours compared to HIV-negative counterparts. Thereby the incidence of malignancies associated with high-risk HPV types is amplified by HIV-induced immunosuppression. Currently available data regarding the effect of antiretroviral therapy on the incidence of HPV-induced neoplasia is conflicting, as some studies have confirmed the expected positive impact by demonstrating decrease in incidence of these tumours following this treatment, and others reporting no such effect. Further examination of HPV co-infection in HIV-infected patients is important not only from the epidemiological perspective, but may provide further insight into host-pathogen interaction in the mechanism of HPV-induced oncogenic transformation.

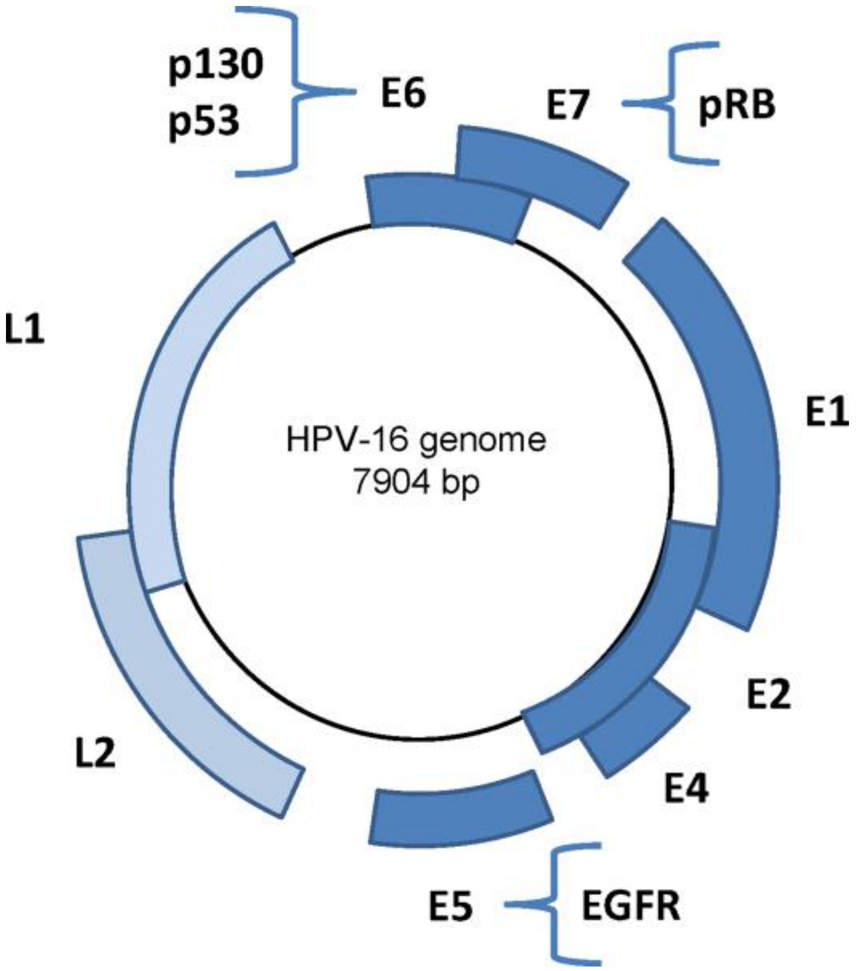


Figure 1.8 The circular genome of HPV-16. The key cellular targets of E proteins are listed: E6 ubiquitinates p53, E7 inhibits pRB, E5 up-regulates epidermal growth factor receptor (EGFR). Figure taken from (Best et al., 2012).

Merkel Cell polyomavirus

The Polyomaviridae family encompasses a group of non-enveloped, icosahedral viruses with a circular, double-stranded DNA genome. The viral DNA is associated with histone proteins of the host cells (H2A, H2B, H3, H4) in a supercoiled fashion, resembling chromatin. The Polyomaviridae are currently divided into three genera and twenty-two different species (Johne et al., 2011), of which eleven are human polyomaviruses described to date.

The first polyomavirus was identified in 1953 as a transmissible agent capable of inducing murine tumours, and was aptly named murine polyomavirus (MPyV) (Spurgeon and Lambert, 2013). In 1960, simian vacuolating virus 40 (SV40) was discovered in rhesus monkey kidney cell cultures. Those cell lines were used in the production of human poliovirus vaccines to combat poliomyelitis, yet at the same time likely infecting large populations with SV40 as a contaminant of the vaccine. The oncogenic potential of SV40, an infectious cause of rodent tumours, desribed later on has lead to a significant concern regarding the administration of a vaccine mixed with an agent capable of neoplastic transformation; however, to date no unequivocal causative link between

SV40 and human cancer has been established (Poulin and DeCaprio, 2006), although the virus has been detected in various human tumours listed in Table 1.6.

Recent technological advances have facilitated rapid discovery of novel viruses. In 2008, a novel DNA polymavirus named Merkel Cell polyomavirus (MCPyV) has been reported in 8 out of 10 Merkel cell carcinoma (MCC) samples (Feng et al., 2008) by a group that has previously discovered Kaposi's sarcoma-associated herpesvirus. MCC is a non-melanoma skin cancer (described in greater detail in Chapter 1.4) associated with immunosuppression, which suggests potential connection with an infectious agent. MCPyV represented the first example of DNA polyomavirus integrating into the host genome, and the initial report confirmed its genomic integration in 6 of the 8 positive MCC samples. In addition, one metastatic sample tested showed an identical clonal pattern to the primary tumour, suggesting causative role of the virus in MCC aetiology.

MCPyV genome contains a conserved replication origin, early genes that code for small and large T antigens, and late genes regions coding for VP1, VP2, and VP3 (Figure 1.6) that encode capsid proteins. Upon infection, early genes are rapidly transcribed, and late genes are activated later in the course of infection to form complete virions. MCPyV genome bears high similarity to the African Green Monkey lymphotropic polyomavirus, and is more distantly related to SV40. Interestingly, a 22-nucleotide viral miRNA (MCV-miR-M1-5p) is transcribed from the late region. The function of this miRNA is perhaps autoregulatory, as it reduces the level of transcripts containing sequences from the early region of the viral genome (Seo et al., 2009).

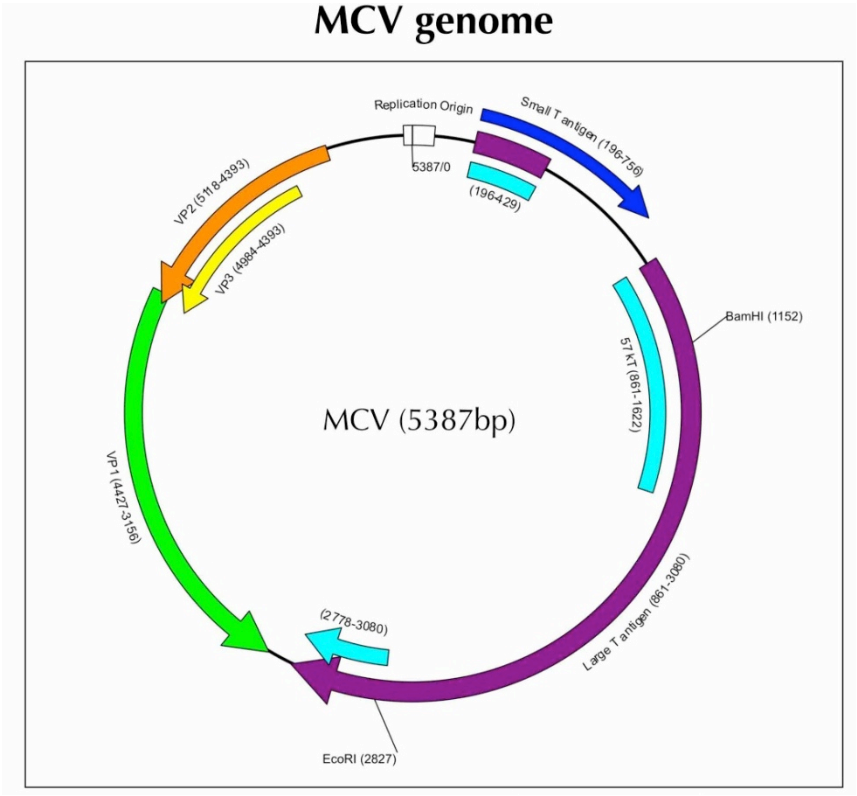


Figure 1.9 Genome of MCPyV. Figure taken from (Feng et al., 2008).

MCPyV Epidemiology

Since its discovery, the presence of MCPyV has been described in various types of lesions and populations: it seems highly prevalent in MCCs originating in European or American patients, but is less abundant in Australian MCC patients (Garneski et al., 2009b). The virus was not detected in neonatal blood spots of 150 children, suggesting the infection does not affect genetic transformation in utero (Gustafsson et al., 2012). The virus was also not detected in proliferative skin disorders (Foulongne et al., 2008), extracutaneous melanoma (Giraud et al., 2008), lymphoid tissue of AIDS patients (Sharp et al., 2009), prostate cancer (Bluemn et al., 2009), or small lung cancer (Wetzels et al., 2009), which morphologically resembles MCC. However, 42% of US sample population were shown to be seropositive for this virus (Kean et al., 2009), and seropositivity increased with advancing age. Specifically, MCPyV seroprevalence is less than 50% among children younger than 10 years of age, 60% in those 10-19 years of age, reaching 81% by the sixth decade of life, and is greater among MCC patients compared to controls (Viscidi et al., 2011). The virus is also highly prevalent among young homosexual men (79%) in which it leads to an asymptomatic infection, and shows no correlation with HIV/AIDS course (Tolstov et al., 2011).

Oncogenic potential of MCPyV

Mutations leading to premature truncation of large T antigen helicase were detected in MCCs positive for MCPyV (Shuda et al., 2008). Among those mutations, multiple pyrimidine dimer substitutions were found, supporting the role of UV light in the onset of MCC upon MCPyV integration. Large T antigen mRNA transcription increases in a dose-dependent manner with UV-exposure (Mogha et al., 2010), further corroborating combined viral and environmental aetiology of MCC.

Recently, it was shown that MCPyV may in fact play a role in chronic lymphocytic leukaemia (CLL). Patients with CLL are at increased risk of MCC, and the virus was detected in 19 CLL samples (Koljonen et al., 2009), 6 of which also possessed novel large T antigen mutations.

A subset of non-small cell lung carcinoma shown to be positive for the virus expressed increased mRNA and protein levels of the BRAF oncogene, while the antiapoptotic BCL-2 gene expression was downregulated in those samples (Lasithiotaki et al., 2013). However, this data was contradicted in a study that detected no significant differences in transcript levels of the K-ras, BRAF, RKIP, p53, RB1, Bcl-2, and Bax genes in between MCPyV positive and negative NSCLC (Hashida et al., 2013).

Molecular data available to date does not indicate clear role of MCPyV in carcinogenesis: although the activation of proteinkinase AKT is detectable in the majority of MCCs, its phosphorylation is unaffected by MCPyV presence, nor by experimental knock-down of the viral large or small T-antigen (Hafner et al., 2012). Moreover, while p53 mutations were detected in MCPyV-positive tumours, those tumours expressed increased pRb levels and were associated with more favourable clinical outcome (Sihto et al., 2011).

A recent study focusing on comparison of cell-cell contacts between Merkel cells and MCC cells (Werling et al., 2011) showed connection of normal Merkel cells to keratinocytes by E- and P-cadherin, desmoglein 2 and desmocollin 2, while the majority of MCCs (90%) contained N-cadherin, and only 67% and 65% contained E- and P-cadherin, respectively. This switch towards N-cadherin is suggestive of epithelial-to-mesenchymal transition, a sign of dedifferentiation and aggressiveness of the tumour. Notably, the observed difference was irrespective of MCPyV infection.

Functional studies assessing the role of MCPyV in MCC cell lines indicate that T antigen is indispensable for MCPyV positive MCC cells maintenance, and that those cells undergo growth arrest or cell death upon T antigen inactivation (Houben et al., 2010). Moreover, this phenomenon was not observed in MCPyV negative cell lines. Those results point towards possible future therapeutic routes, and additionally suggest that the aetiological background of MCC is multifactorial, and MCPyV positive and negative tumours may in fact represent distinct entities with shared histological features.

MCPyV and clinical course of MCC

Antibodies to large and small T antigen of MCPyV were shown to be relatively specific for MCC-affected patients, and the rise in their titre seems to predict spread of the disease (Paulson et al., 2010). The actual presence of the virus in tumour specimen does not seem to affect clinical course of the disease (Handschel et al., 2010, Becker et al., 2009, Garneski et al., 2009b), although other studies suggest that patients infected with MCPyV show prolonged remission (Laude et al., 2010) and better clinical outcome (Bhatia et al., 2010, Sihto et al., 2011). An immunohistochemical investigation has shown that the absence of MCPyV is associated with p53 and KIT expression, however this seemed to represent a negative prognostic feature (Waltari et al., 2010). The precise role of MCPyV in cancer onset and progression remains to be elucidated.

Other human pathogens and cancer

While virus-induced oncogenesis has received considerable attention during the past century, the role of bacteria and parasites in cancer formation has been clarified to a lesser degree. The human body is estimated to be colonised by some 10^{14} microbial cells (Savage, 1977), and 500 different bacterial species are present in the colon alone (Berg, 1996). While the vast majority of human bacteria are considered harmless or even symbiotic (for example, certain intestinal bacteria synthesise vitamin K2), epidemiological evidence has implicated bacteria as likely culprits in the onset of certain cancers. The mechanisms through which bacteria induce cancer include detrimental alterations in host physiology that lead to chronic inflammation, antigen-driven lymphoproliferation, increasing the levels of signalling molecules that induce proliferation of epithelial cells, and direct mutagenic effects of their toxins (Chang and Parsonnet, 2010).

Helicobacter pylori is a common human pathogen affecting 50% of the world population, and a well-established causative agent of gastric cancer. It leads to 60% of all stomach cancers world-wide, in addition to countless precancerous lesions (chronic and atrophic gastritis, intestinal metaplasia and dysplasia). It is the cause of gastric adenocarcinoma, the most common malignant tumour of the stomach associated with poor prognosis, and gastric mucosa-associated lymphoid tissue (MALT) lymphoma, a less common type of gastric neoplasia which can be practically cured by H. pylori eradication in majority of cases (Stathis et al., 2009).

Other bacteria and parasites associated with cancer are listed in Table 1.7

Pathogen	Associated Cancer	Comment
Opisthorchis viverrini	Cholangiocarcinoma	Food-borne trematode parasite (liver fluke) endemic to Southeast Asia
Schistosoma haematobium	Bladder cancer	Parasitic flatworm (blood fluke) highly prevalent in Sub-saharan Africa
Chlamydia pneumonia	Lung cancer	Bacteria associated with increased risk of lung cancer
Streptococcus bovis	Colon carcinoma	Opportunistic pathogen; bacteraemia strongly associated with colorectal tumours
Salmonella typhi	Gallbladder Cancer	Highly prevalent in developing countries

Table 1.7 Bacteria and parasites epidemiologically associated with cancer.

1.2. Anatomy and biology of the skin

The skin is a dynamic, resilient organ that acts as the frontline barrier against external noxa harmful to the human organism. Specifically, it guards the body against the entry of microorganisms, solar radiation damage and excessive fluid loss. In addition to this protective role, the skin serves as a sensory organ perceiving external sensations through various receptors, regulates the body temperature, modulates the immune system by allergen recognition, and serves as the site of vitamin D synthesis.

Skin is the body's largest organ (Wysocki, 1999), with the average area of an adult's skin of 1.7 m^2, four kilograms of weight, and approximately eleven miles of blood vessels. Additional skin-associated anatomical structures enhance its particular functions: hair and hair follicles contribute to the skin's protective, thermoregulatory and sensory role, sebaceous glands produce lipids and sebum for lubrication and protection, sweat glands aid in thermoregulation, and nails mechanically protect the distant phalanges. In order to cope with its many functions, the skin and its appendages have developed a complex differentiation and regeneration process that constantly renews this adaptive yet highly resistant integument (Alonso and Fuchs, 2003).

Skin histology

The skin is composed of three basic components that are of different embryologic origin, and serve different functions: the epidermis derived from the ectoderm, the dermis and subcutaneous tissue (subcutis, hypodermis) which are of mesodermal origin.

Subcutis functions as an anchor linking skin to structures below (bones, muscles), and predominantly consists of adipocytes, fibroblasts and macrophages, elastin and loose connective tissue. It contains nerves and larger blood vessels, and due to the presence of many fat cells, it serves as an insulation layer for the body. On top of the subcutis lies the corium, or dermis.

The dermis is composed of fibroblasts, macrophages, adipocytes, collagen, elastin, glycosaminoglycans. It contains additional anatomical structures such as blood and lymph vessels, nerve endings, hair follicles, sweat and sebaceous glands. Dermis and epidermis are connected by a basement membrane, which consists of extracellular matrix and growth factors, and provides a stable anchoring for epidermal keratinocytes.

The epidermis, sometimes designated as interfollicular epidermis, is the uppermost layer of the skin comprised of stratified squamous epithelium. It consists of keratinocytes (the principal epidermal components), pigment cells (melanocytes), somatosensory receptors (Merkel cells) both derived from the neural crest, and antigen-presenting Langerhans cells.

Histologically, the epidermis is divided into additional five layers (Figure 1.6): stratum basale is the deepest layer separated from the dermis by the basement membrane. This layer contains a pool of cells with proliferative potential (stem cells and their immediate progeny, the transient amplifying cells), the descendants of which undergo terminal differentiation once they leave this compartment (Iglesias-Bartolome et al., 2013). Thus, stratum spinosum contains mitotically inactive differentiated keratinocytes which produce keratohyalin granules as they enter stratum granulosum, in addition to cornified envelope proteins. Once cellular organelles are destroyed, the differentiation process culminates and results in a cornified envelope surrounding each dead, flattened, transcriptionally inactive keratinocyte (corneocyte). Stratum lucidum, present only in areas of thick skin such as the palms of hands, thereby consists of dead, fully keratinised anuclear cells, and is rich in protein-bound lipids and eleidin, a keratohyalin derivative. Stratum corneum is the outmost layer composed of multilayered corneocytes that are shed off through gradual desquamation. Maintaining the equilibrium between keratinocyte renewal and differentiation is crucial for epidermal homeostasis, during wound repair and adaptation to environmental challenges (Fuchs, 2009).

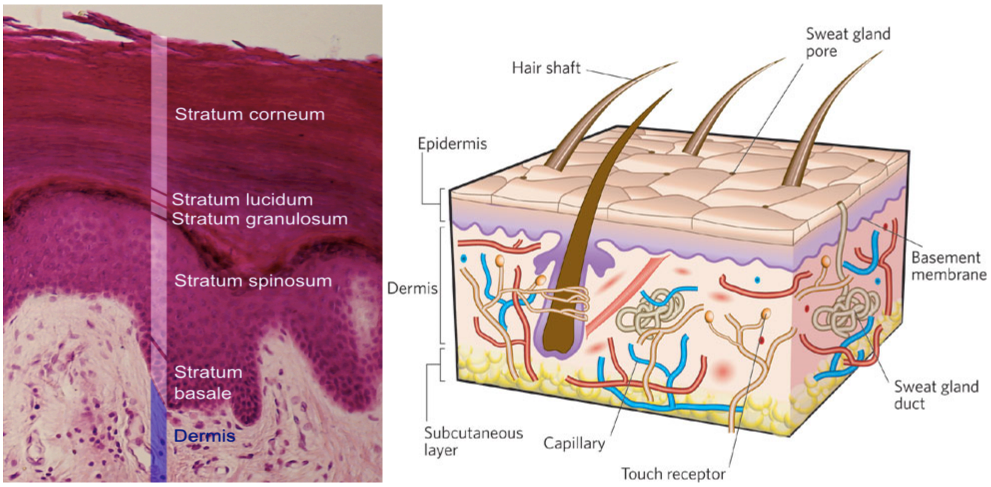


Figure 1.10 Structure of the human skin (left) and the epidermis (right). Figure taken from (MacNeil, 2007) and http://en.wikipedia.org/wiki/File:Epidermal_layers.png.

Homeostasis of the epidermis

In order to maintain its protective and regulatory function, the epidermis is under tight homeostatic regulation: periodically, proliferative basal keratinocytes enter the programme of terminal differentiation, which leads to their unidirectional, upwards movement alongside loss of proliferative potential, keratinisation, cornification and eventual cell death (Fuchs, 2008). The key step of the process is the transition from stratum basale to stratum spinosum of the epidermis, during which cells entering stratum spinosum undergo profound transcriptional, morphological and functional changes. Specifically, basal keratinocytes exit from the cell cycle, switch off the expression of keratin 5 (KRT5) and KRT14, markers of epithelial cells with proliferative potential, and turn on the expression of KRT1 and KRT10 and additional structural proteins (S100). These form a complex system of connections interlinked with cell junctions, leading to increased mechanical strength and resistance. The entry into this process is regulated by p63, MYC and the canonical Notch signalling pathway (Koster et al., 2004, Blanpain et al., 2006, Gandarillas and Watt, 1997).

During later phases of terminal differentiation, filaggrin precursor profilaggrin accumulates in keratinocytic granules. Other proteins such as loricrin, involucrin and trichohyalin are synthesised and deposited under the plasma membrane and become crosslinked by transaminases, generating a highly resistant envelope, which co-creates a waterproof lipid bilayer between individual cells. When the differentiation process is completed, dead keratinocytes are wrapped in hydrophobic cornified envelope, and evenetually shed off, while being replaced by differentiating cells from the basal layer (Blanpain and Fuchs, 2009).

Transcriptional regulation of keratinocyte differentiation

Although our knowledge of transcriptional control during epidermal differentiation is far from complete, several molecules were implicated in driving epidermal lineage commitment. The depletion of apoptosis-stimulating protein of p53 (iASPP), an oncoprotein and p63 co-regulator, leads to terminal stratification of keratinocytes (Chikh et al., 2011) and iASPP functions as inhibitor of senescence in epidermal homeostasis (Notari et al., 2011).

FOXN1 (WHN) is transcribed early during the differentiation process, where it induces KRT1 expression while suppressing profilaggrin, loricrin, and involucrin, safeguarding the gradual character of the process (Baxter and Brissette, 2002). Additionally, FOXN1 is important for nail and hair follicle homeostasis (Potter et al., 2011).

Homeobox gene Distal less-3, DLX3, is expressed in the suprabasal layers of the epidermis at the onset of differentiation, and following activation by p63, it reduces the proliferative potential of basal keratinocytes (Di Costanzo et al., 2009).

Other important regulators include the POU family of transcription factors (OCT1, OCT6, OCT11) that activate various differentiation genes (Andersen et al., 1997), KLF4 that promotes epidermal differentiation and maintenance (Sen et al., 2012), and OVOL1 that reduces the proliferative potential of keratinocytes entering terminal differentiation (Nair et al., 2006).

MiR-203 is an evolutionarily conserved microRNA highly expressed in the epidermis, in which it inhibits the replicative potential and promotes differentiation through targeting p63 (Lena et al., 2008, Yi et al., 2008). MiR-203 is capable of inducing cell cycle arrest rapidly upon its activation (Jackson et al., 2013), and in the absence of miR-203, cell proliferation is elevated even in suprabasal layers of the skin (Yi et al., 2008).

Additional epigenetic regulators of terminal differentiation in the epidermis include ING5, SMARCA5, BPTF, EZH2 and UHRF1, the depletion of which drives keratinocytes towards terminal differentiation. This seems to be the result of a complex interplay between these genes and their targets, including P63, ITGB6 and epidermal stem cell marker ITGB1 (Mulder et al., 2012).

Histone modifications have also been implicated in the modulation of regulators that control epidermal differentiation. Tri-methylation of lysine 27 on histone H3 (H3K27me3) marks are enriched in promoters of a subset of genes involved in epidermal differentiation in undifferentiated keratinocytes, and are lost during calcium-induced differentiation (Sen et al., 2008). Setd8 histone methyltransferase was shown to play an essential role in histone modifications at H4, required for the survival of epidermal stem cells and in cooperation with MYC, for epidermal differentiation (Driskell et al., 2012). Epidermal stem cells are discussed in greater detail in section 1.6.

1.3. Non melanoma skin cancer: an overview

1.3.1. Human skin cancer

Skin cancer is the most common human malignancy (Harris et al., 2001). Based on the cell of origin, skin cancer can be divided into two major categories: melanoma and non-melanoma skin cancer (NMSC). As this nomenclature suggests, the first type of skin cancer is caused by malignant transformation of melanocytes, and the second by neoplastic transformation of all other types of skin cells. Melanoma is less common than NMSC, but is associated with significant mortality, leading to more than 2000 deaths a year in the UK only (Neal and Hoskin, 2009).

Over the past decades, the incidence of NMSC has substantially increased, predominantly in Caucasian populations (Rodust et al., 2009), representing now the most common malignancy in this ethnic group (Trakatelli et al., 2007). The term NMSC encompasses a diverse cluster of malignancies and includes cutaneous lymphoma, adnexal tumours, sarcomas, Merkel-cell carcinoma, but mainly basal cell carcinoma (BCC) and squamous cell carcinoma (cSCC) (Madan et al., 2010). Approximately 75% of NMSCs are BCCs, and about 20-25%% are cSCCs, which makes cSCC the second most common malignancy in humans.

The overall mortality from non-melanoma skin cancer is rather low compared with melanoma and other solid tumours (Clayman et al., 2005), but varies across different types of NMSC. MCC has a 5-year survival of just 60% (Albores-Saavedra et al., 2010a), while BCC practically never forms metastases (Wong et al., 2003) and is associated with excellent prognosis. However, in contrast to the vast majority of BCCs, squamous cell carcinoma has a significant propensity to metastasise (Cherpelis et al., 2002). Moreover, both tumours are associated with considerable morbidity due to local invasion and disfigurement (Corona, 1996) (Figure 1.6). This can substantially impact patients' psychosocial health and wellbeing, since very commonly those lesions are located in sun-exposed, and thus usually uncovered areas of the body (such as the head, neck and upper extremities), and are frequently multiple.

Figure 1.11 Panel of clinical photos showing disfigurement following treatment of NMSC.

Although the prognosis of cSCC patients is generally favorable (three-year disease specific survival is 100% if a patient has no associated risk factors (Clayman et al., 2005)), because of its immense prevalence, cSCC represents a serious burden to national health-care systems (Housman et al., 2003). In addition to this, the prognosis of patients with metastatic disease is dismal, with less than 20% ten-year survival for patients with disease present in regional lymph nodes (Claerhout et al., 2010), and less than 10% for patients with identifiable distant metastases (Madan et al., 2010). In recessive dystrophic epidermolysis bullosa (RDEB) patients, a rare autosomal recessive blistering disorder caused by mutation in type VII collagen (Dunnill et al., 1994, Hovnanian et al., 1992), the risk of developing an cSCC is especially high, and cSCC represents the most common cause of premature death in RDEB patients (Fine et al., 2008). Fifty-five percent of Hallopeau–Siemens RDEB patients, the most severe form of the disease, die from metastatic cSCC by 40 years of age (Pourreyron et al., 2007). Thus it is of imminent importance to unweave the biological process of cSCC carcinogenesis, namely its propensities to metastasise, in order to provide better therapeutic perspectives for affected patients.

Figure 1.12 Clinical presentation of RDEB patients skin. Figures taken from (Fine, 2010).

1.3.2. Clinical features of NMSC

1.3.2.1. Basal cell carcinoma

BCC is the most common malignancy diagnosed in humans, affecting millions of people every year worldwide. This tumour develops most commonly on sun-exposed areas (80% of BCCs develop in head and neck area (de Vries et al., 2005)), most likely due to accumulated UV damage. Additionally, the disease is associated with Gorlin syndrome; a rare autosomal-dominant disease characterised by defects of the skin, eyes, bones and other body systems, in which numerous BCCs appear bilaterally and in areas not exposed to UV-light. The underlying cause of the Gorlin syndrome is a mutation in tumour suppressor gene PTCH (Devi et al., 2013).

Macroscopically, BCCs present as pearly, red or pink plaques and nodules, which may ulcerate. Although this tumour rarely metastasises, it is capable of local invasion and can lead to severe disfigurement. The most common treatment is surgical excision, but radiotherapy, topical imiquimod cream and a range of other destructive modalities such as cryotherapy and photodynamic therapy are also used. Over the past five years, targeted therapies for the rare cases of metastatic or otherwise untreatable BCC have been developed, based on the inhibiton of the PTCH tumours suppressor pathway, frequently upregulated in BCC (Goppner and Leverkus, 2011, Tabs and Avci, 2004, Von Hoff et al., 2009).

While the highly conserved PTCH pathway plays a crucial role in many processes during embryonic development, such as proliferation and differentiation (Merchant and Matsui, 2010), it is inactive in the majority of adult tissues. However, stem cells, hair follicles, and skin cells rely on the pathway for maintenance (Cowey, 2013). PTCH is a 12-pass transmembrane receptor, and Hedgehog (Hh) ligand initiates the pathway signalling by binding to PTCH. Once coupled with Hh, PTCH lifts its inhibitory effect on Smoothened (Smo), a G-coupled receptor. Smo then propagates the signal by uninhibiting glioma-associated protein (Gli) through the suppressor of fused molecule (Sufu), promoting its oncogenic effects through a variety of target genes involved in cell cycle, apoptosis resistance and angiogenesis (Cowey, 2013). Thereby, PTCH functions as a tumour suppressor. Other solid tumours shown to harbour abnormal hedgehog pathway signalling include medulloblastoma and rhabdomyosarcoma (Merchant and Matsui, 2010).

A variety of anti-Smo agents were tested and incorporated in clinical research, and vismodegib, an orally available inhibitor of Smo, has received FDA drug approval in 2012. The drug demonstrated 30-50% overall response rate among patients with metastatic BCC (Cowey, 2013), and significantly reduced the number new of BCCs in patients with Gorlin syndrome compared to placebo (Tang et al., 2012). Although effective in many patients with locally advanced and metastatic BCC, eventually most tumours have developed resistance to vismodegib and progressed. Nonetheless, Smo inhibitors represent a clinically relevant treatment modality for patients with advanced disease and Gorlin syndrome, highlighting the translational potential of molecular analysis in cancer treatment.

1.3.2.2. Squamous cell carcinoma

Cutaneous SCC (cSCC) is the second most common skin cancer among Caucasians (Alam and Ratner, 2001). In spite of identified environmental factors involved in cSCC development and increasing public awareness of these, the incidence of cSCC continues to rise (Madan et al., 2010). Cutaneous SCCs usually present as hyperkeratotic nodules which are often ulcerated and painful. Based on histology, the disease is divided into morphological categories which reflect the degree of differentiation within the tumour: well differentiated (WD) cSCCs are eosinophilic, showing striking similarity to normal keratinocytes, and display a pattern of layers not unlike normal squamous epithelium. Moderately differentiated (MD) cSCC is less similar to normal squamous epithelium and the typical architecture of epithelial layers is less defined. Poorly differentiated (PD) cSCC lacks epithelial architecture and tumour cells show a high degree of morphological atypia. Representative haematoxylin-eosin stained slides are shown in Figure 1.9.

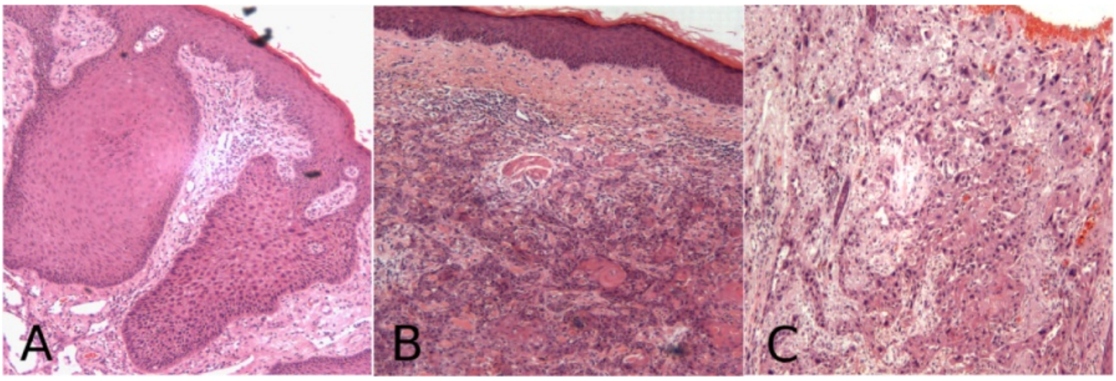


Figure 1.13 Panel of histological specimen representing three histological degrees of cSCC differentiation. A – WD, B – MD and C – PD SCC (4x magnification).

A proportion of cSCC will behave aggressively and metastasise; cSCC accounts for 20% of skin cancer-related deaths, and poor outcome is seen in 16% of tumours with a thickness greater than 6.0 mm (Brantsch et al., 2008).

Unlike BCC, precursor lesions for cSCC are recognised. Skin SCCs frequently arise from cutaneous areas bearing multiple actinic keratoses (AK), generally regarded as precursor lesions for cSCC. These lesions are common, affecting about 23% of the UK population over the age of 60 years (Harvey et al., 1996). They usually appear as asymptomatic red, scaly papules on sun-exposed areas (Berman and Cockerell, 2013), with histological evidence of epidermal dysplasia, and are more common among Caucasian individuals of advanced age. So-called “field cancerisation” is an area of the skin (often chronically sun-exposed) in which multiple AK occur (Braathen et al., 2007). Field cancerisation can often be a relatively broad area with no obvious macroscopic changes in which pro-neoplastic mutations are already present, and thus represents the preneoplastic stage of cSCC development (Rubin, 2011b). Although the precise mechanism of field cancerisation development has not been entirely clarified, molecular data acquired from tumours of the head and neck, esophagus, and bladder indicate that a single genetic event occurs in an individual cells, due to clonal expansion then spreads laterally, replacing normal epithelium and creates above-described preneoplastic fields (Dakubo et al., 2007).

Cumulative UV radiation exposure is the main culprit in AK onset (Schmitt and Miot, 2012). First, it induces the initial genetic mutations in skin cells, and then contributes to the growth of transformed lesions. UV damage generally leads to apoptosis resistance, genomic alterations, inflammatory, and immunosuppressive disruptions which in the epidermal context result in keratinocyte transformation, aberrant proliferation and eventual clinical appearance of AKs. Sunscreen application significantly reduces the number of newly-formed AKs in a dose-dependent manner, and may lead to the involution of existing lesions (Thompson et al., 1993). Additionally, the impact of seasonal variation on AK diagnosis has been noted, with the peak number of newly diagnosed AKs registered during spring months (Hancox et al., 2004). This trend may be the reflection of seasonal UV radiation variability and sun exposure. Because AKs can progress to malignant cSCC, clinically detected AK must be treated (Braathen et al., 2007).

There is significant controversy over the rate of progression of an individual AK to cSCC; a systematic review of 15 studies found progression rates between 0.025% and 20% per year per lesion (Quaedvlieg et al., 2006). About 20-25% of AK will regress over the course of a year, although the majority will later reappear, and numbers of prevalent AKs are strongly influenced by ongoing sun-exposure. A recent study showed that, although the risk of progression for a specific AK was only 2.57% at 4 years, 65% of all primary cSCC arose directly in lesions previously diagnosed as AK (Criscione et al., 2009). The remainder of cSCC are presumed to arise from sub-clinical AKs.

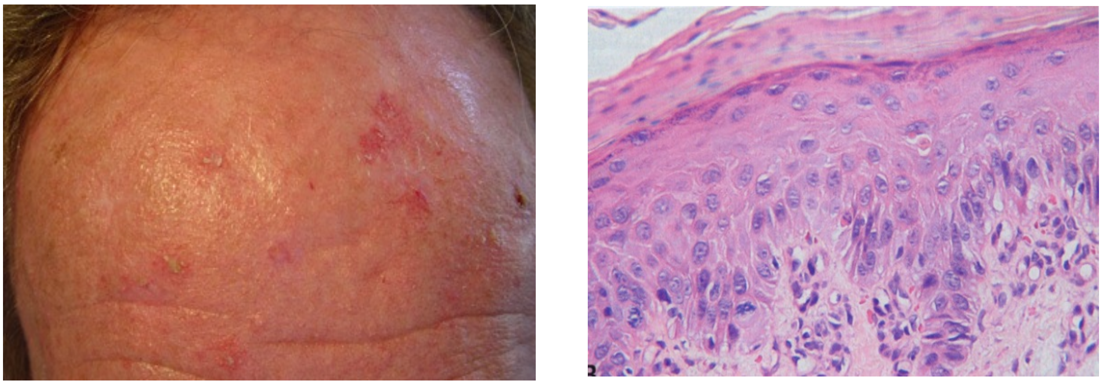


Figure 1.14 Clinical and microscropic images of AK.

The epidemiology, risk factors and molecular basis of cSCC are described in more detail in section 1.4.

1.3.2.3. Merkel Cell carcinoma

Merkel Cell carcinoma (MCC) is a rare primary neuroendocrine tumour of the skin most commonly diagnosed in elderly people. It was first described by Cyril Toker in 1972 as a trabecular carcinoma of the skin based on morphological observations (Toker, 1972). During the past decades, the incidence of MCC has increased considerably (Hodgson, 2005), and it is estimated to affect approximately 400 people a year in the UK (Albores-Saavedra et al., 2010b). The tumours occur most frequently on sun-exposed sites, and are increasingly common with advancing age, with 95% of MCC diagnosed in people over the age of sixty (Albores-Saavedra et al., 2010b). This malignancy is more common in white population compared to dark-skin people (Miller and Rabkin, 1999), and is also more prevalent in men than in women.

In solid organ transplant patients, the incidence of MCC is increased 10-40 fold compared to the general population, affected patients are generally younger at the onset (50% are under 50 years of age) (Stoff et al., 2010), and 8% of all MCC occur in OTRs (Buell et al., 2002).

Clinically, MCC is an aggressive skin malignancy, usually presenting as a rapidly enlarging, firm, reddish nodule on a sun-exposed site. This tumour is locally invasive and often metastasises, and in spite of multidisciplinary treatment approach is associated with a significant mortality (almost one third of affected patients die from the disease within two years (Schwartz et al., 2013).

Figure 1.15 Clinical presentation of MCC on scalp. Left panel shows recurrent tumour.

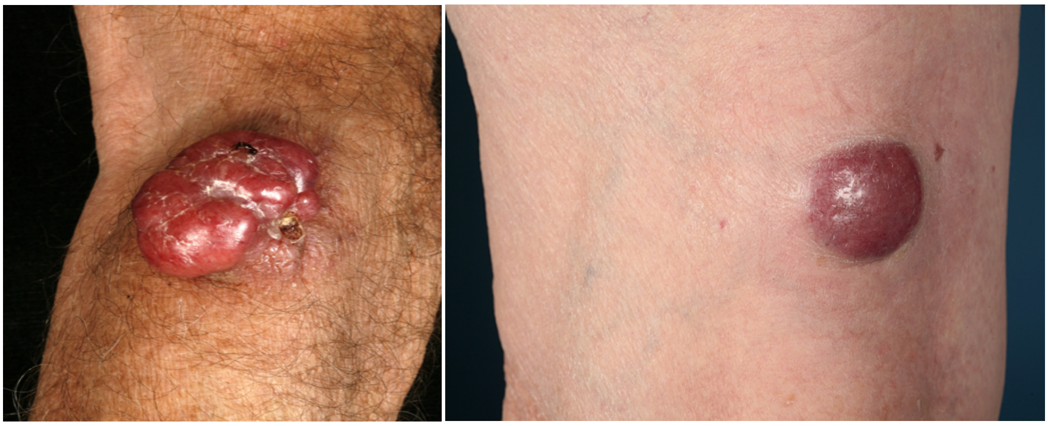


Figure 1.16 Clinical presentation of MCC of the arm in two different patients.

MCC was originally thought to arise from Merkel cells, which were first described in 1875. In the cytoplasm of these cells, electron-dense Merkel granules are found using electron microscopy, and such granules are seen in the cytoplasm of MCC cells, which lead to the postulation that Merkel cells are the cell of origin for this malignancy. However, this theory has not been conclusively proven so far.

1.3.2.4. Other NMSCs

Other NMSCs include cutaneous T-cell lymphoma, a disease with peak incidence among populations of thirty-fifty years of age, that often clinically resembles eczema or other benign skin lesions. B-cell non-Hodgkin's lymphoma also affects the skin, albeit rarely, and is associated with a favourable prognosis if diagnosed early. Kaposi's sarcoma is a malignancy of vascular origin mostly associated with acquired immune deficiency syndrome; however, it can occur sporadically in endemic populations and immunosuppressed patients. The lesions are usually purple plaques or nodules that may be ulcerated or bleed (Figure 1.11).

Other rare cutanoues tumours include sarcomas such as dermatofibrosarcoma protuberans, angiosarcoma and liposarcoma; appendageal/adnexal tumours such as hidradenocarcinoma, cylindroma and sebaceous carcinoma, and histiocytoses such as Langerhans cell and non-Langerhans cell histiocytosis.

Figure 1.17 Kaposi's sarcoma of the shin. Kaposi's sarcoma is a common cutaneous malignancy in AIDS-patients associated with HHV-8. Figure taken from (Mehta et al., 2011).

1.4. Epidemiology and pathogenesis of cutaneous squamous cell carcinoma

1.4.1. Epidemiology of cSCC

Skin cancer is currently considered to have reached epidemic proportions (Salasche, 2000). Because of prominent differences between national registries in data collection and processing, precise figures on cSCC incidence are difficult to obtain. However, a sharp incline in cSCC incidence has been observed during the past decades in many areas across the globe and suggests that cSCC is an imminent health-problem world-wide. In the UK, there are an estimated 30,000 new cases per year, and there has been an approximate doubling in incidence over the past 15 years. The increase in recorded cases may be due to various factors, including increasing UVR exposure and population aging, since cSCC is predominantly a disease of older people. CSCC is also numerically and economically a major burden to the NHS and other public health-care systems.

1.4.2. Risk Factors for cSCC development

Individual risk of cSCC development depends upon various genetic, phenotypic and environmental factors. The key factors include skin tone, accumulated sun damage to the skin, radiation exposure and compromised immune system.

1.4.2.1. UVR exposure

Various lines of evidence corroborate the causative role of ambient solar radiation in cSCC carcinogenesis. CSCC occurs mainly on sun-exposed body sites, and its occurrence can be effectively prevented by sun protection. It is also well-established that geographical location is associated with different trends in cSCC incidence; the relative risk of cSCC is three times as high among people born in areas with high amounts of UV radiation as among people who move to those areas during adulthood (Alam and Ratner, 2001). Sun-sensitive skin phenotype provides further evidence of the role of UVR in that cSCC risk is two to five times as high in people with fair skin, hazel or blue eyes and blonde or red hair compared to those with darker features. Thus, it is now accepted that exposure to UVR is the most common risk factor for cSCC; UVB radiation is principally responsible for the tumour-inducing process with UVA having an additive role in the oncogenic mechanism: UVB causes direct damage to RNA and DNA by inducing covalent bond formation between adjacent pyrimidines, leading to mutagenic products (cyclopyrimidine dimers [(G:C to A:T)], pyrimidine-pyrimidine adducts), while UVA causes indirect damage through the formation of reactive oxygen species and by causing DNA damage in deeper strata of the skin compared to UVB. A thirty-year follow-up study of psoriasis patients treated with psoralen and UVA has described a greatly increased risk of cSCC in those patients (Stern, 2012). Additionally, border wavelengths between UVA and UVB of around 315 nm are especially mutagenic to the skin (Ikehata et al., 2013).

In addition to an unequivocal mutagenic effect, UVB radiation has an immunosuppressive potential in skin (Phan et al., 2006). It induces the production of immunosuppressive cytokines, and profound changes in antigen-presenting cells (Xu and Elmets, 2012). Following a single low dose of UVB radiation, Langerhans cells, antigen-presenting cells of the skin and mucosa, are reduced by 50% (Murphy et al., 1993), yet epidermal Langerhans cells are essential for the generation of UVB-induced regulatory T lymphocytes that mediate local immune reaction. Additionally, UVB radiation leads to profound transcriptome changes in the skin, including upregulation of genes that encode antimicrobial peptides, inflammatory molecules, and genes that play a role in activation of innate defense and early adaptive immune pathways (Kennedy Crispin et al., 2013).

1.4.2.2. Ionising radiation

Ionising radiation has also been implicated in the development of cSCC. Originally it was suggested that exposure to therapeutic radiation is associated with the development of BCC, but not cSCC. Nonetheless, it was subsequently demonstrated that the risk of cSCC may be increased with radiotherapy, especially in individuals prone to sunburn and particularly if experienced by individuals younger than 20 years of age.

1.4.2.3. Human papillomavirus (HPV) infection

HPV infection has also been associated with cutaneous SCC. The first link was observed in patients with epidermodysplasia verruciformis (McGregor and Proby, 1996, Orth, 1986) (Figure 1.12), a rare inherited skin disorder, who are prone to beta-HPV infection and up to 60% of those develop cSCC, which is in most cases HPV-positive.

Based on PCR testing of tumours from immunocompetent (IC) patients with cSCC, 30-50% of these are positive for HPV, while immunosuppressed (IS) patients show 65-90% HPV positivity in their cSCCs (Harwood and Proby, 2002). However, a more sensitive technique using reverse hybridisation shows almost equal distribution of HPV infection in both IC and IS patients (Purdie et al., 2009). Within a month of transplantation, 80% of patients are seropositive for at least one HPV type, mostly beta-HPV, and the majority retain this serostatus (Antonsson et al., 2013). Interestingly, sun exposure of more than four hours a day is associated with increased HPV seropositivity among transplant patients (Sampogna et al., 2012). In addition, the risk of cSCC is associated with seropositivity for beta-HPV types (Bavinck et al., 2010, Karagas et al., 2006, Proby et al., 2011). For the actual cSCC risk, high viral load may be a more important factor than infection with a specific HPV type (Neale et al., 2013).

While thus far no clear hierarchy of specific HPV types in cSCC has been conclusively defined, there is a trend towards beta types oncogenic in EV, notably HPV-5, -8, -36 and -38 (Casabonne et al., 2007), which fits well with functional studies showing additional pro-carcinogenic activities for HPV8 and -38 E6 and E7 proteins (Storey and Simmonds, 2009). On the other hand, high-throughput sequencing of cSCC transcriptome has failed to detect HPV expression in any cutaneous SCC sample (Arron et al., 2011, Ganzenmueller et al., 2012). The rather striking variability of HPV findings suggests that the role of HPV may be indirect in the pathogenic process, potentially by facilitating UV-related carcinogenesis by prevention of UV-induced apoptosis or impairing DNA repair (Madan et al., 2010).

Figure 1.18 Clinical presentation of Epidermodysplasia verruciformis.

1.4.2.4. Immunosuppression

Cutaneous SCC is the most common post-transplant malignancy in Caucasian organ transplant recipients (Reichrath and Nurnberg, 2008). The risk of cSCC increases 65 to 250 fold in comparison with non-transplant population (Madan et al., 2010). CSCC occurs more frequently in organ transplant recipients compared to BCC, while in non-transplant population the ratio of BCC and cSCC is reversed. Additionally, the biological characteristics of cutaneous SCC are more aggressive in OTR patients in comparison to the general population: cSCCs generally develop at younger age, starting three to five years after transplantation, demonstrating high frequency of local recurrence (13.4%) during the first 6 months after excision and a high frequency of lymph node metastasis (7%) during the second year after excision (Reichrath and Nurnberg, 2008). Moreover, cSCCs in OTRs grow rapidly to a greater size (>2 cm diameter) and often demonstrate an aggressive histological growth pattern, being frequently associated with perineural invasion or invasion of the cartilage, fat, or bone (Reichrath and Nurnberg, 2008). Furthermore, cSCCs in patients with diseases affecting the immune system, such as HIV/AIDS or non-Hodgkin lymphoma, also show an aggressive profile (Madan et al., 2010). It is of interest that transplant patients with a history of non-melanoma skin cancer prior to transplantation are at an increased risk of metastatic non-melanoma skin cancer post transplantation, and that transplant patients with cSCC are at higher risk of internal malignancies compared to OTR free of this skin cancer (Wisgerhof et al., 2012). The overall incidence of malignancy in transplant patients ranges from 7-15%, thus the majority of transplant patients remain cancer free, yet those patients who develop skin malignancies are also at increased risk of other tumours. This suggests potential intrinsic genetic factors are “unmasked” by immunosuppression, increasing the overall susceptibility to cancer.

1.4.2.5. Other risk factors

Smoking is another risk factor for the development of cSCC. While the evidence associating smoking with basal cell carcinoma and melanoma remains controversial, the association between smoking and cSCC has been confirmed by several studies (De Hertog et al., 2001, Grodstein et al., 1995, Aubry and MacGibbon, 1985, Karagas et al., 1992, Lear et al., 1998). Other risk factors include exposure to arsenic (Petter and Haustein, 2000), coal tar, insecticides, herbicides and petroleum products (Gallagher et al., 1996), and high-fat diet, particularly in people with previous history of skin cancer (Ibiebele et al., 2007).

1.4.3. Molecular genetic basis of SCC

1.4.3.1. Cytogenetic abnormalities

Chromosomal aberrations are frequently observed in both precursors (90% of actinic keratoses) and neoplastic skin lesions (95% of cSCCs) (Ashton et al., 2005). Aneuploidy rates are estimated to reach 25-80% in cSCC (Carless and Griffiths, 2008), with a substantial difference in aneuploidy rates between well (46%) and moderately (75%) differentiated cSCC (Pilch et al., 1994). Aneuploidy rates for AK have been estimated at 69%, and 89-92% for cSCC in situ, which is higher than those reported for both cutaneous cSCC and AK, suggesting that DNA aneuploidy may not be a good prognostic marker of cutaneous cSCC (Ashton et al., 2005).

A variety of methods have been used to investigate the cytogenetic alterations associated with cSCC. In one study, karyotypic analysis of 13 short-term cultured primary tumours together with 10 previously published cSCC cases detected recurrent numerical gains, of which the most frequent was the loss of chromosome 21 found in 41% of cases (Jin et al., 1999). These data are comparable to karyotypic findings in other squamous cell carcinomas, such as SCC of the head and neck, breast, and lung (Ashton et al., 2005).

Fluorescence in situ hybridisation (FISH) of interphase chromosomes highlighted 3p21 deletion as a recurrent aberration in cSCC (Dobler et al., 1999), while polymorphic microsatellite marker analysis identified 9p, 13q and 17 as other frequent targets of deletion (Quinn et al., 1994). Metaphase comparative genomic hybridisation (CGH) was used in three studies to investigate the genetic changes associated with cSCC. Overall, frequent recurrent gains of genetic material at 3q, 8q, 17q and 20q and losses at 3p, 4q9p, and 18q were observed, with each study also identifying further recurrent changes specific to their subset of cSCCs (Ashton et al., 2005, Carless and Griffiths, 2008).

More recently, higher resolution array-based techniques have been used to more accurately map the genetic events to particular chromosomal regions. One such study (Purdie et al., 2009) used single nucleotide polymorphism (SNP) microarrays to map the genomic instability in 60 tumours, revealing a previously unidentified mechanism of loss of heterozygosity (LOH) in cSCC known as uniparental disomy (UPD). UPD is a phenomenon whereby mitotic non-dysjunction or mitotic recombination causes the loss of one allele and subsequent duplication of the remaining allele, resulting in LOH without loss of genetic material (copy number neutral LOH). The most frequent aberrations identified in this study were LOH at 3p and 9p. In contrast, another study using high resolution array CGH to analyse 16 cSCC identified gain at 3q as the most common event (Salgado et al., 2010). Genome-wide SNP analysis of 13 cSCCs collected from transplant patients detected various chromosomal aberrations in 6 cSCC samples, with 9p21 loss as the most common (Hameetman et al., 2013).

Many of the gross chromosomal aberrations associated with skin carcinogenesis are merely a consequence of genomic instability and may be irrelevant to tumour progression. In addition to large-scale chromosomal gains and losses, high resolution microarray-based techniques are able to detect focal deletions and gains and may therefore facilitate the identification of genes critical for cSCC onset.

1.4.3.2. Specific genetic abnormalities in cSCC

The above-mentioned SNP microarray analysis of 60 SCC (Purdie et al., 2009) identified 9 tumours with microdeletions at 9p23 within the locus of the candidate tumour suppressor gene, protein tyrosine phosphatase receptor type D (PTPRD; (Sato et al., 2005, Stallings et al., 2006)), which were associated with greater aggressiveness of tested tumours. Deletions of PTPRD were subsequently shown to be associated with metastatic cSCC, and somatic mutations of the gene were found in 37% of cSCC tumours (Lambert et al., 2012).

Deletions at the CDKN2A locus at 9p21.2-9p21.3 were detected in both studies using SNP arrays (Purdie et al., 2009, Hameetman et al., 2013), alongside deletions within the Fragile histidine triad (FHIT) gene at 3p14.2, both of which have been previously associated with cSCC development (Kubo et al., 1997, Brown et al., 2004).

Array-comparative genomic hybridisation (aCGH) identified a focal amplification at the locus of the CKS1B gene at 1q21.1-q21.3 as a marker of aggressive clinical behavior in cSCC (Salgado et al., 2010). CKS1B is a member of the highly conserved cyclin kinase subunit 1 (CKS1) protein family whose amplification has previously been shown to correlate with poor prognosis in multiple myeloma (Chang et al., 2006). Another aCGH study has revealed the loss of 17p.13.3-17q25.3, where p53 is located, in almost half of tested cSCC samples (Li et al., 2012), and gains of 7p21 and 7q31 corresponding to SKTS5 in 10% of samples (Fleming et al., 2013).

Whole-exome sequencing of 11 cSCCs has revealed mutations in Notch genes in majority of these tumours, mostly consistent with UV damage. These mutations were cSCC specific, as whole-exome sequencing of five BCC samples did not identify mutations in either NOTCH1 or NOTCH2. In cSCC, there was no association with histological subtype or p53 mutations (Wang et al., 2011).

Other studies have used a candidate gene approach to investigate tumour suppressor genes and oncogenes implicated in carcinogenesis in other organs. UV signature p53 mutations have been detected in an overwhelming majority of cSCC, but are also present in actinic keratoses and sun-exposed normal skin, and thus would appear to be an early event in UVB-induced skin carcinogenesis (Benjamin and Ananthaswamy, 2007). Mutational activation of the *ras* oncogene is a characteristic event in human carcinogenesis, but its role in cSCC remains controversial. Considerable variation exists in the reported frequency of activating *ras* mutations in these tumours, ranging from 67% of cSCC from a group of patients with the rare genetic disorder xeroderma pigmentosum (XP, (Daya-Grosjean et al., 1993)) to 0% of tumours in another cohort of XP patients (Sato et al., 1994) and from 0% to approximately 20% in the general population. However, studies among melanoma patients treated with RAF inhibitors that seem to develop keratoacanthomas and cSCC has shown a higher frequency of activating RAS mutations among patients treated with these agents (Oberholzer et al., 2012, Su et al., 2012b).

Focal amplification of the c-myc oncogene is another recurrent event in human carcinogenesis which has been implicated in cSCC development. To date, this aberration has been identified in a high proportion of tumours, ranging from 50% in a series of renal transplant recipient cSCCs (Pelisson et al., 1996) to 63% in the general population (Toll et al., 2009).

One study investigated the genetic basis behind the reduced expression of thrombospondin-1 (TSP-1) in cSCC, an anti-angiogenic matrix glycoprotein. FISH analysis of normal skin and 5 cSCCs revealed that loss of expression correlated with loss of one copy of chromosome 15, the location of TSP-1 (Burnworth et al., 2007).

Tetraspanin CD151 was shown to be elevated in human cSCC compared to other cutaneous tumours, and seems to regulate keratinocyte proliferation through interaction with STAT3 and integrins (Li et al., 2013b), suggesting it may be a potential player in early cSCC carcinogenesis.

The apparent genetic heterogeneity and overall lack of consensus on the role of specific genes in cSCC tumourigenesis and progression suggests that additional molecular mechanisms including epigenetic modifications may drive the onset of this malignancy.

1.5. Epigenetics and cancer

1.5.1. Overview

In 1942, Conrad Waddington originated the term ‘epigenetics’ to describe the interaction between the genome and the environment, which ultimately determines the cellular phenotype (Dawson and Kouzarides, 2012). Although cells of the same organism have practically identical DNA that remains generally constant throughout the cell’s lifetime, the physiological functions and properties of cells vary greatly and even change on a single cell level. This dynamic regulation is partly due to the genomic and epigenomic interplay.

Epigenetics as a scientific field represents the exploration of heritable changes that are not mediated at the DNA sequence level (Cheung and Lau, 2005). Epigenetic regulation is a relatively highly conserved mechanism that is capable of underpinning different phenotypic expression of an organism in spite of the inherited DNA code: for example, mice in Figure 1.19 are identical in their DNA sequence, yet their coat colour varies due to epigenetic variation at the agouti locus in their genome.

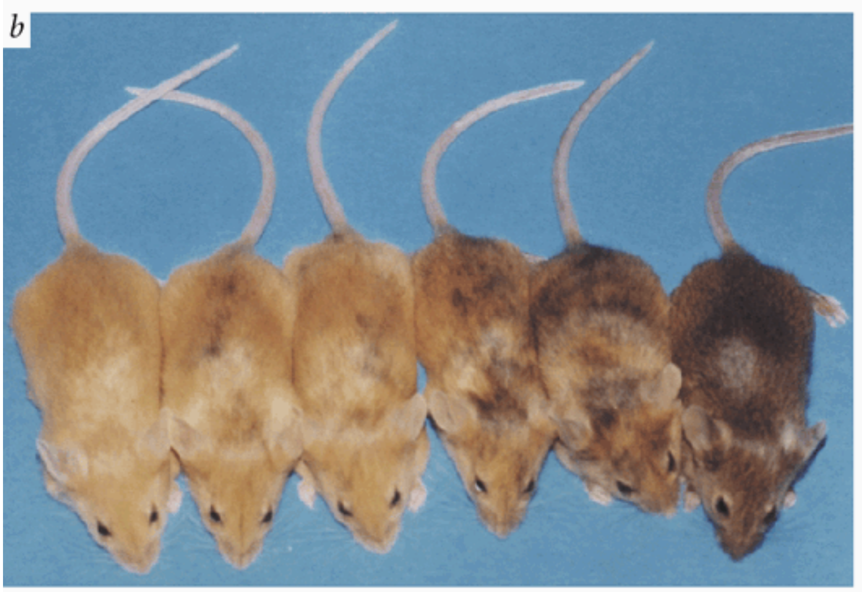


Figure 1.19 Variation in the fur coat colour due to epigenetic alteration. All depicted mice are genetically identical. Figure taken from (Morgan et al., 1999).

There are several molecular mechanisms that mediate epigenetic regulation, and include namely DNA methylation and chromatin or histone modifications (Cheung and Lau, 2005), RNA interference (RNAi) and prions (Yool and Edmunds, 1998). Several other important 'epigenetic factors' have also been described to play a role in many processes, such as intracardiac hemodynamic forces during heart and the cardiac conduction system development (Hove et al., 2003, Reckova et al., 2003), but the central mechanisms currently understood as epigenetic regulators are DNA methylation, histone modification and RNAi.

Epigenetic regulation has been linked to several physiological processes, such as embryonic development (Reik et al., 2001), control of the circadian rhythm (Masri and Sassone-Corsi, 2010), placental development (Maltepe et al., 2010), ageing (Rakyan et al., 2010, Gronniger et al., 2010), as well as diseases: cancer (Esteller, 2008), obesity, depression, cardiovascular diseases (Chen et al., 2011), neurodegenerative prion diseases (Tatzelt and Schatzl, 2007) and many more.

From the perspective of a therapeutic intervention, drugs affecting or altering epigenetic regulation already have an established role in the clinical setting: DNA methyltransferase (DNMT) inhibitor 5-aza-2′-deoxycytidine (decitabine) and 5-azacytidine is currently approved for clinical application in haematological cancers (Piekarz and Bates, 2009), suberoylanilide hydroxamic acid (vorinostat) and romidespin are histone deacetylases (HDAC) inhibitors currently approved for clinical application in cutaneous T-cell lymphoma patients (Zain et al., 2010). Many more agents acting on the epigenetic level of regulation have been assessed or are currently being tested in preclinical trials (Helin and Dhanak, 2013).

All above-mentioned factors corroborate the attractiveness of studying epigenetic mechanisms in both organic processes and disease. During the past decade, epigenetics have become a highly influential field of research, with important discoveries spanning diverse areas of research: from epidemiology (Heijmans et al., 2008), through environmental sciences (Kundakovic et al., 2013) to evolutionary paterns of commercially important invertebrate (Nanty et al., 2011). Linking epigenetic modifications with cellular functions provides novel insight into multiple levels of regulation, and novel targets for therapeutic intervention.

1.5.2. DNA methylation

One of key epigenetic mechanisms acting in higher organisms is DNA methylation (Deeb et al., 2010). This process occurs on DNA level and involves the addition of methyl group to 5' carbon of the cytosine ring (Figure 1.14). This process is mediated by methyl-transferase enzymes. The genome-wide distribution of methylated cytosines represents an additional level of regulatory information that affects cellular phenotype in addition to the DNA sequence. While DNA methylation patterns can be transmitted from one generation to another, it is still flexible enough to adapt for different somatic cell types, or environmental impulses (Laird, 2003).

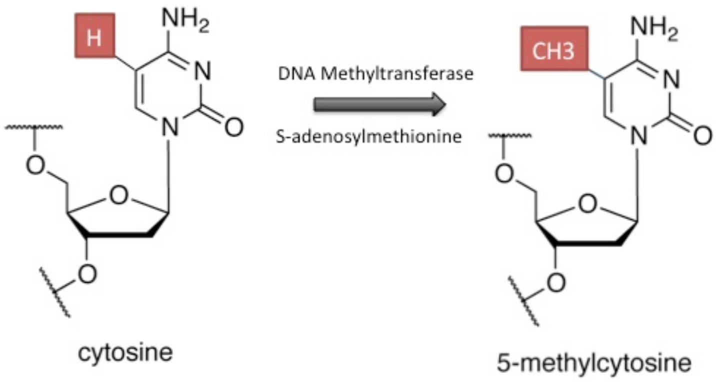


Figure 1.20 The enzymatic process of C5’ methylation. Methylation of the 5’-carbon in the cytosine ring is carried out by methyltransferase enzymes. S-adenosylmethionine serves as the donor of the methyl group.

The process of methylation occurs in vertebrates most commonly at CpG sites: regions found in the genome which are rich in a symmetrical cytosine-guanine sequence, since methyl-transferases recognise 5'-CpG-3' motif (Laird, 2003). However, non-CpG methylation that targets CpA, CpT, and CpC dinucleotides (Ziller et al., 2011) has been shown to occur in both embryonic stem cells (Ramsahoye et al., 2000) and to a lesser extent, somatic cells (Ziller et al., 2011). CpG regions usually exhibit lower level of methylation in comparison with the rest of the genome, which is normally more methylated.

Methylation often occurs in the promoter region of a given gene, upstream from the transcriptional start site, since about 60% of CpG islands in the human genome are found within promoter regions (Antequera, 2003). Several proteins such as MeCP2, MBD1, MBD2, MBD3 and MBD4 share the methyl-CpG-binding domain (MBD), which is a specific structure capable of binding to methylated CpG (Ballestar and Wolffe, 2001). Upon binding of MBD proteins, several transcription repressors including HDACs are recruited to the CpG site and transcription of the gene is thus blocked (Jones and Takai, 2001). Abnormal methylation can, for example, thereby lead to the suppression of tumour-suppressor genes (Zhang et al., 2006b), or conversely, aberrant demethylation of an oncogene promoter can be linked to cancer initiation (Koslowski et al., 2011). CpG methylation is absent in mitochondrial DNA (Hong et al., 2013).

1.5.3. Non-CpG methylation

Although described as potentionally important regulatory mechanism in mammalian cells over two decades ago (Woodcock et al., 1987), and as pivotal in embryonic stem cells over a decade ago (Ramsahoye et al., 2000), non-CpG methylation remains largely underexplored in most organisms. It has been so far mostly studied in plants, but recent studies show 20-25% non-CpG methylation in stem cells, namely in CpA context. The observed non-CpG methylation disappears with differentiation, indicating that non-CpG methylation is involved in the maintenance of the pluripotent state (Lister et al., 2009). In non-small cell lung carcinoma, non-CpG methylation of p53 exon 5 is more prevalent compared to normal lung tissue (Kouidou et al., 2005). In Burkitt’s lymphoma, overall non-CpG methylation was shown to be generally low, but significantly enriched at non-CpG sites within genes (Kreck et al., 2013).

Recent data evaluating non-CpG methylation across various mammalian tissues has estimated the overall prevalence of non-CpG methylation at approximately 7.5% of all non-CpG cytosines (Yan et al., 2011). Additionally, high levels of non-CpG methylation occur in oocytes, but not in sperm, and change dramatically during early post-conception development (Tomizawa et al., 2011). In post-developmental cells, non-CpG methylation is much more common in neurons, which are non-dividing cells (Kozlenkov et al., 2013), suggesting its role cannot be exclusively linked to pluripotency maintenance, but more likely linked to tight regulation of specific cellular processes.

1.5.4. Hydroxymethylation

In spite of early hypotheses that considered DNA methylation to represent a relatively stable epigenetic mark, it became gradually clear that it is a highly dynamic modification that is either added or removed through an enzymatic process (Dawson and Kouzarides, 2012).

A set of so-called TET (ten-eleven translocation) proteins with mammalian DNA hydroxylase properties has been shown to catalytically convert 5’-methylcytosine to 5’-hydroxymethylcytosine (5’hmC) using molecular oxygen (Thomson et al., 2013, Tahiliani et al., 2009). Further oxidation of 5’hmC into 5-formylcytosine and 5-carboxylcytosine creates nucleoside intermediates that are eventually excised through the mechanism of base-excision repair, resulting in DNA demethylation (Thomson et al., 2013).

Current line of evidence suggest an important role of hydroxymethylation in preimplantation embryos (Inoue and Zhang, 2011), DNA methylation fidelity control (Williams et al., 2011), mouse embryonic stem cells maintenance through Nanog expression (Ito et al., 2010), differentiation (Ficz et al., 2011) and transcriptional regulation in mouse embryonic stem cells (Wu et al., 2011, Xu et al., 2011). High levels of 5'hmC were detected in human cerebellar neurons (Kriaucionis and Heintz, 2009), while the loss of of this epigenetic mark was described in melanoma epigenome (Lian et al., 2012) and in Alzheimer's disease (Chouliaras et al., 2013). Given that various hematopoietic malignancies are clearly linked to TET-mediated oncogenesis (Moran-Crusio et al., 2011, Lorsbach et al., 2003, Figueroa et al., 2010a), and changes in 5'hmC in response to carcinogens were described in the liver (Thomson et al., 2013), it seems plausible that a congruent interplay of mechanisms regulating DNA methylation plays a role in both physiological regulation and disease.

1.5.5. MicroRNAs

The classically translated genes are not the only molecules capable of regulating gene expression in the cell. MicroRNAs (miRNA) are non-coding single stranded RNA molecules, approximately 22 nucleotides in length, capable of regulating gene expression by binding to the 3' untranslated region (UTR) of messenger RNA (mRNA) and thus repressing translation. Each miRNA can target multiple genes and it is estimated that the expression of up to a third of all human protein-coding genes is possibly regulated by miRNAs (Esquela-Kerscher and Slack, 2006). Additionally, miRNA can accelerate target mRNA degradation and thus indirectly reduce the eventual amount of translated protein (Shyu et al., 2008).

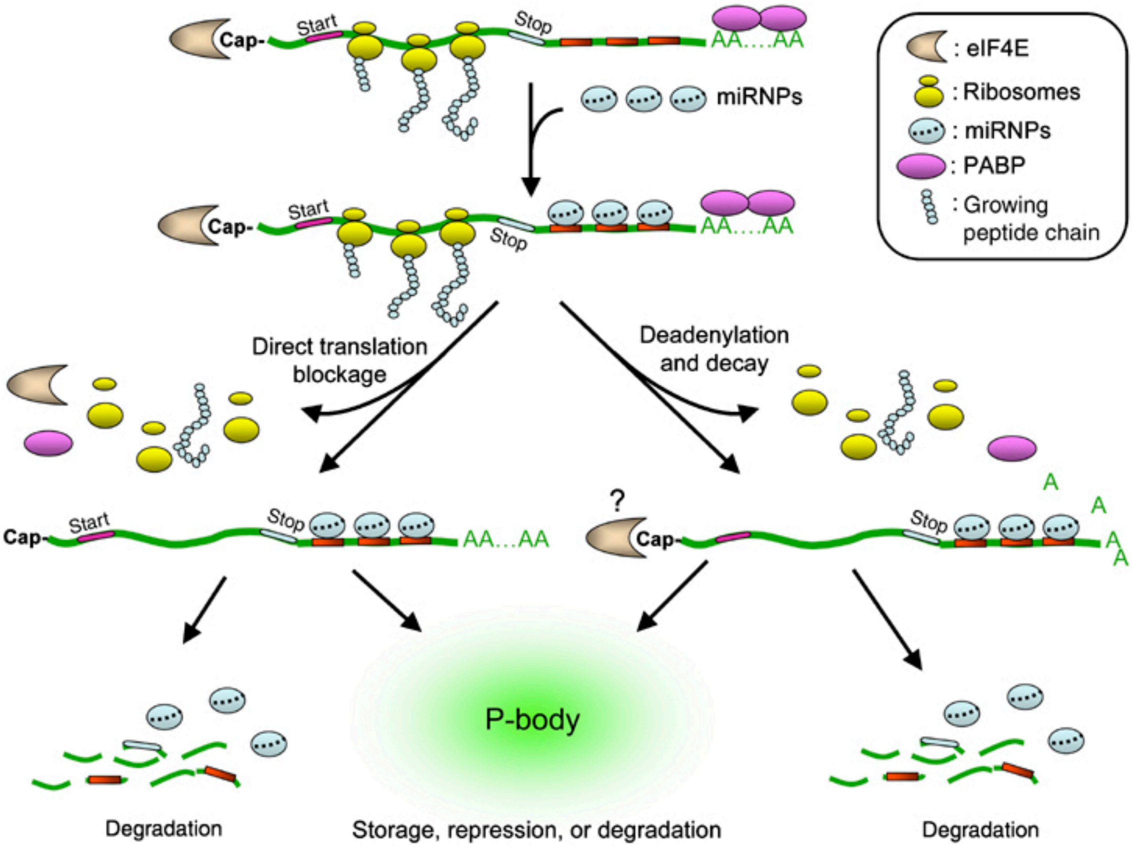


Figure 1.21 Basic processing of miRNAs. miRNP=microRNA ribonucleoprotein complex, PABP=poly(A)-binding protein. Figure taken from (Shyu et al., 2008).

MiRNAs are incorporated into RNA-induced silencing complex (RISC), and then guide effector complexes (miRNPs) to the target mRNA site through hybridisation with targtet mRNA 3'UTR sequence, leading to direct inhibition of protein translation.MiRNPs are also able to deadenylate target mRNA and thus accelerate its degradation. See Figure 1.21.

MiRNAs are capable of affecting many fundamental cellular processes in both animals and plants (Shyu et al., 2008), and have been linked to a range of diseases including schizophrenia, cardiomyopathy and cancer. Examples of tumour suppressor miRNAs include the miR-15a and miR-16-1, shown to be frequently deleted in B-cell chronic lymphocytic leukaemia and leading to overexpression of the anti-apoptotic protein B-cell lymphoma 2 (Bcl-2) (Calin et al., 2002, Cimmino et al., 2005). MiR-21 has an anti-apoptotic function and is upregulated in glioblastomas and breast cancer (Chan et al., 2005, Iorio et al., 2005).

In skin pathology, miR-99a that regulates the expression of Insulin-like growth factor 1(IGF-1R) has been found to be downregulated in psoriasis (Lerman et al., 2011), and hematopoietic-specific miR-142-3p has been found in psoriatic lesions (Joyce et al., 2011), suggesting a complex regulatory interplay in the genesis of this lesion. MiRNA dysregulation in skin cancer is described in greater detail in section 1.5.7.

1.5.6. Histone modification

Histones allow the packaging of DNA within the nucleus. The complex of DNA and histones is called a nucleosome, and is formed by the octamer of histones H2A, H2B, H3 and H4 in pairs. The formation of histones is mediated by so-called histone-chaperone proteins, and this process is essential for genome stability and preservation of epigenetic information following DNA replication, repair and gene transcription (Burgess and Zhang, 2013). Approximately 146bp of DNA are wrapped around the core of histones to form a nucleosome. This complex allows dense packaging of DNA and is called chromatin. Chromatin is a highly dynamic cellular structure critically important in regulating transcription, DNA replication, repair, and recombination, and any alterations to chromatin structure can lead to severe perturbations of those processes (English et al., 2006).

Heterochromatin, considered to represent a transcriptionally inactive form of DNA packaging, is often found in segments of repetitive DNA elements, including rRNA genes, centromeres, telomeres and interspersed transposable elements (Bierhoff et al., 2013). Euchromatin, on the other hand, is considered transcriptionally active, although it has gradually become clear that this distinction is not sufficient to cover chromatin complexity: For example, heterochromatin becomes transcriptionally permissive during S-phase of the cell cycle, and its relative active transcription strongly correlates with developmental stages of tissue (He et al., 2012).

Histone modification is important in many processes such as DNA repair, mitosis and apoptosis and histones are heavily modified post-translation. The N-terminal end of the histone protrudes from the complex, containing about 28% of aminoacids of the histone core, and is a subject to various modifications: acetylation, methylation, sumoylation, ubiquitination, phosphorylation, namely at lysine residues. Given that both the aminoacid sequence of histones and the character of the regulatory marks is evolutionarily conserved, it seems plausible that these are critical for the maintenance of cellular functions in eukaryotes (Thompson et al., 2013).

The acetylation of positively charged lysine residues leads to a local opening of the chromatin, which allows transcription factors and chromatin remodelling factors to access it and ultimately results in gene transcription. The process of acetylation is facilitated by histone acetyltransferases (HATs) and conversely removed by histone deacetylases (HDACs) (Zhao et al., 2013b). Histone methylation occurs on arginine or lysine residues and can lead to both transcriptional activation and suppression (Cloos et al., 2008). Histone methylation depends on the activity of histone methyltransferases and is removed by histone demethylases. Perturbations of histone posttranslational modifications have been associated with cancer.

One of the most explored histone modifications includes trimethylated lysine 4 of histone 3 (H3K4me3). This epigenetic mark is induced by the Trithorax group of proteins (Coskun et al., 2012), and has been associated with practically all active gene promoters. Although the role of this modification has not been entirely clarified yet, because not all genes whose promoters bear it are transcribed, it is presumed to render gene “poised for activation”, possibly due to promoter demethylation (Bernstein et al., 2006).

Conversely, H3K27me3 is found in promoters of transcriptionally silenced genes, and H3K27me3 pre-marked genes are targeted by DNA methylation (Severson et al., 2013). This mark is induced by the Polycomb group complex and has been linked to silencing genes important in embryonic development. Irrespective of other epigenetic marks present in the gene promoter, if H3K27me3 is present, the gene is always silenced (Wang et al., 2008).

Trimethylation of histone H3 on Lys36 (H3K36me3) is induced by SETD2 trimethyltransferase (Li et al., 2013a). It is preferentially enriched in exones compared to introns, and may play a role in alternative splicing, because its levels are lower in alternatively spliced exones (Kolasinska-Zwierz et al., 2009). Possibly, it may play a role in transcription elongation and might prevent transcription from cryptic start sites within the coding region (Carvalho et al., 2013).

The H3K36me3 mark has recently been implicated in human DNA mismatch repair, because it recruits recognition protein hMutS-α onto chromatin. Cells that lack SETD2 show microsatellite instability and an increased rate of spontaneous mutations, which is typical for cells deficient in DNA mismatch repair (Li et al., 2013a), indicating that perturbations in this modification may be linked to cancer susceptibility (Li, 2013).

The role of histone modification in embryonic stem cells and pluripotency regulation is described in detail in Chapter 1.6.

1.5.7. Epigenetics and non-melanoma skin cancer

From the non-melanoma skin cancer perspective, previous epigenetic studies have focused mainly on DNA methylation of cSCC, BCC and Merkel Cell Carcinoma (MCC). Studies assessing methylation in Merkel Cell carcinoma, a rare yet very aggressive form of skin cancer, are limited to date. A study using methylation-specific polymerase chain reaction (PCR) has revealed methylation of p14ARF promoter in 42%, and p16INK4a methylation in a single sample (Toyoshima et al., 2008). Another study using the same technique but a substantially higher number of samples (21 vs. 98) found methylation of RASSF1A gene and CDKN2A in 42 and 9 MCC samples, respectively. Other genes tested (TP73, PTPRG and FHIT) were not found to be methylated (Helmbold et al., 2009a). The same group looked at RASSF1A and p16 gene promoter methylation to find out that those genes are methylated in 94% and 56% of samples tested, respectively (Helmbold et al., 2009b).

Methylation analysis of p16, a putative tumour suppressor, in basal cell carcinoma detected methylation in 7/28 tested samples (Soufir et al., 1999) using methylation restriction analysis. Another study looking at methylation at different histological areas of BCC showed that p16 methylation diminished in tumour cells that were located at the invasive front (Svensson Mansson et al., 2007), underscoring the role of histological heterogeneity in molecular studies of cancer.

The promoter region of another tumour suppressor, T-cadherin, has been found to be methylated in 6 of 25 BCC samples using methylation-specific PCR (Takeuchi et al., 2002b). The methylation status of Gli1, an oncogene that is overexpressed in BCC, was not altered in this tumour (Ghali et al., 1999).

14-3-3sigma is a cell cycle inhibitor regulated by p53. A study assessing methylation status of 14-3-3sigma in 41 BCC samples using methylation-specific PCR showed that 28 (68.3%) samples are hypermethylated in the gene promoter region and in 22 cases this was concordant with reduced protein expression (Lodygin et al., 2003). Elsewhere, a study of methylation of a panel of selected genes in 68 BCC samples using the same technique has shown that LAMA3, LAMC2, CDH1 and CDH3 genes are methylated in more than 30% of the samples, LAMB3, P16, RASSF1A, CDH13, CDH1, CDH3 and beta IG-H3 were methylated in more than 0% but less than 30% of the samples, while Cyclin D2 (CYCD2), TJP2a and CLDN7 were not methylated in any sample tested. This study also found a difference in methylation of LAMA3 and CDH1 (p value = 0.02 and 0.04, respectively) when they stratified for sun-exposure of the samples tested; however, given that the number of samples from sun-exposed sites was 61 and only 7 from sun-protected areas, the validity of this testing is limited (Sathyanarayana et al., 2007). Fragile histidine triad (FHIT) gene is a pro-apoptotic molecule and its inactivation plays a role in cellular DNA damage survival. Methylation-specific polymerase chain reaction and combined bisulphite-dependent restriction analysis (COBRA) of this gene methylation status in 17 BCCs has shown that in 15/17 of samples this gene is hypermethylated, and this was confirmed by COBRA which reported methylation in 88% of the samples (Goldberg et al., 2006). The decreased expression of FHIT was confirmed on the protein level using immunohistochemistry.

The Patched (PTCH) gene is an important regulator during embryonic development, but also functions as a tumour suppressor. Mutations in this gene are associated with Gorlin syndrome, and since its mutation in sporadic BCC has been reported, a few studies have attempted to evaluate the methylation status of PTCH in BCC. A study using methylation-specific PCR found no difference in PTCH promoter methylation using samples treated with RNALater or culture media for an unspecified period of time (Cretnik et al., 2007), while another study has shown that the source material is a key factor in methylation studies in BCC: five of 16 BCCs showed methylation of PTCH when fresh frozen samples were used, while only two of 16 samples showed methylation of the gene when identical paraffin-embedded samples were used (Heitzer et al., 2010) and high-resolution melting technology was applied.

Methylation-specific high-resolution melting (MS-HRM) analysis of AKAP12 gene in 85 BCC samples showed 0% methylation in 11, and 10% methylation in 74 of BCC samples, while 78 of adjacent normal tissue samples showed 0% and 7 of them were methylated in 10% degree. Although this difference was significant in terms of statistical difference (p value < 0.01), the biological significance of this difference is questionable, because 10% degree of methylation difference may not be sufficient to lead to actual biological difference in cell regulation.

Methylation studies in cSCC have so far identified several potential targets of aberrant methylation, including putative tumour-suppressor genes. T-cadherin promoter methylation was detected in 12 out of 28 invasive squamous cell carcinomas (Takeuchi et al., 2002a). A study assessing promoter hypermethylation of the E-cadherin in 9 cases of spongiotic dermatitis (control), 9 cases of actinic keratosis, 8 cases of squamous cell carcinoma in situ, and 7 cases of invasive squamous cell carcinoma gene detected E-cadherin hypermethylation in 6 of 7 cases (85%) of invasive squamous cell carcinoma, 4 of 8 cases (50%) of squamous cell carcinoma in situ, 4 of 9 cases (44%) of actinic keratosis, and 2 of 9 cases (22%) of control, suggesting that this gene promoter hypermethylation increases with more malignant profile of cutaneous lesions.

Using the same sample panel, death-associated protein kinase promoter methylation was found in 1 squamous cell carcinoma in situ and 1 control. P16 promoter hypermethylation was detected in 1 invasive squamous cell carcinoma and 1 control only, suggesting no difference in methylation between cSCC and control. However, a subsequent study exploring methylation of p16 in 36 cSCC cases found 13 (36%) samples to be aberrantly hypermethylated (Brown et al., 2004). In addition, this study found hypermethylation of p14 in 16 of 38 (42%) cSCC cases. In recessive dystrophic epidermolysis bullosa cSCC, two out of eight cSCC samples exhibited p16 methylation (Arbiser et al., 2004).

In skin cancer, microarray profiling of mycosis fungoides (MF) and benign inflammatory conditions showed 49 miRNAs to be differentially expressed at the tumour stage MF compared to benign inflammatory dermatoses. The most significant differentially expressed were miR-155 and miR-92a and both were upregulated (van Kester et al., 2011). Mir-211, which has tumour suppressor properties, is silenced in melanoma (Levy et al., 2010), mir-125b is down-regulated in metastatic melanoma (Glud et al., 2010), and mir-21 is up-regulated in Sezary syndrome (van der Fits et al., 2011).

Basal cell carcinoma profiling has identified an array of dysregulated miRNAs, including let-7 involved in cell proliferation, UV-induced miR-21 upregulation (Heffelfinger et al., 2012), miR-17 regulated by the MAPK/ERK axis (Sand et al., 2012b) and other miRNAs involved in critical cellular processes.

In cSCC, miRNA profile is affected by UV radiation. Mir-21 and mir-184 expression is significantly increased in SCC, while mir-203 expression was shown to be downregulated in this tumour (Dziunycz et al., 2010), as well as the expression of miR-124 and miR-214 which regulate ERK1 and ERK2, respectively (Yamane et al., 2013).

These data support the role of epigenetic changes in non-melanoma skin cancer, and justify an integrative exploration of various epigenetic mechanisms in this very common cancer.

1.6. Stem Cells

In recent years, stem cells have attracted wide-spread interest not only from the scientific community, but also from a large number of media outlets and the general public. This considerable public attention largely owns to the proclaimed immense potential and promise of stem cells in tissue repair and regenerative biology (Pera and Trounson, 2004), thereby conceivably spelling a new era in human medicine. In mammals, one single zygote gives rise to over 200 types of adult cells (Boyer et al., 2005), and exploiting this property for therapeutic purposes is extremely tempting. Nevertheless, while raising hope among many patients, the question of potential clinical application of stem cells derived from human embryos has ignited a fierce ethical controversy, even leading to a presidential veto to a bill allowing federal funding for embryonic stem cell research in the USA (Dolgin, 2006). However, stem cells are attractive not only from the perspective of tissue engineering and cellular therapy, but also in the study of embryogenesis, developmental biology, disease modelling and cancer research.

Defining stem cells

Stem cells (SC) are essential for normal embryonic development and homeostasis of an organism, since SC play a fundamental role in the initiation and progress of prenatal growth, subsequent organogenesis and in the maintenance and regeneration of mature tissues. The definition of stem cells is commonly based on their capacity of self-renewal and concurrent differentiation towards multiple types of tissues. Mammalian ontogenesis is characterised by gradual restriction of cellular developmental potential, and thus determines the “stemness” of cells at various stages of embryogenesis: *totipotent* (or omnipotent) stem cells have the potential to create an entire organism, and arise upon egg fertilisation and zygote formation. Totipotency is maintained up to the eight-cell stage of the morula. Once the cells reach the blastocyst stage, consisting of outer trophoblast cells (trophoectoderm) and undifferentiated inner cell mass (ICM), totipotency is lost, and the cells of the ICM are *pluripotent*; they possess the capacity to generate all three primary germ layers: the endoderm, mesoderm, and ectoderm. *Multipotent* stem cells are even more limited in terms of tissues those can generate; usually within the tissue lineage[3]. The hierarchy of stem cells is shown in Figure 1.22.

---

[3] Unipotent stem cells are by definition precursor cells.

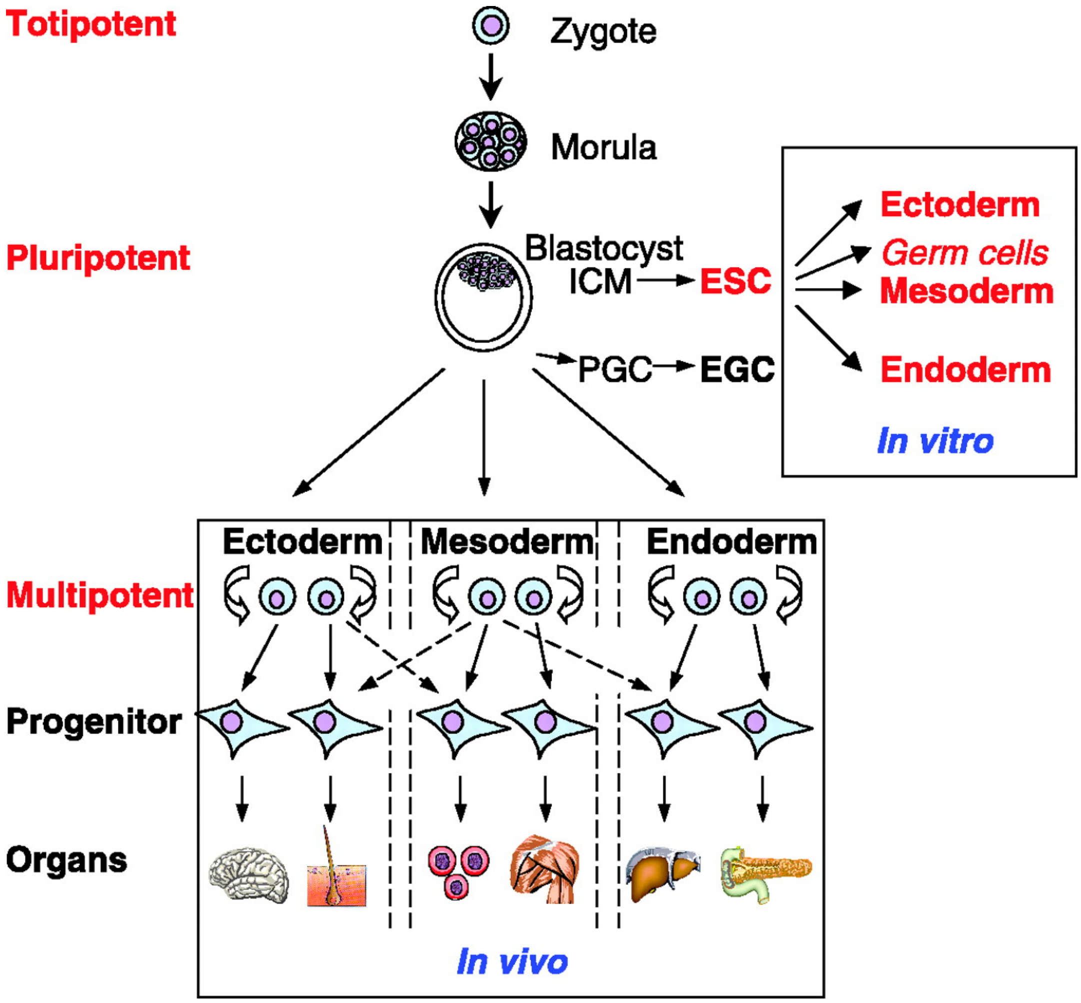


Figure 1.22 Hierarchy of stem cells. Figure taken from (Wobus and Boheler, 2005).

1.6.1. Embryonic stem cells

Embryonic stem cells (ESCs) are derived from the blastocyst, which represents the pre-implantation stage of mammalian embryos. As described above, ICM contains the pluripotent ESCs, which under physiological conditions develop into the actual organism, and the surrounding trophoblast is determined to become a part of the placental chorion.

Historically, pluripotency was first demonstrated in 1964 in embryonal carcinoma cells, when upon in vivo implantation, single cell grafts formed various clonal lines including well differentiated somatic tissues (Kleinsmith and Pierce, 1964). In 1981, two research groups independently demonstrated pluripotent cells derived from early mouse embryos that were expanded in vitro while retaining their pluripotency (Evans and Kaufman, 1981, Martin, 1981). Upon insertion into murine blastocyst, these cells contributed to the embryogenesis and germ line formation, and this ability of murine ESCs lead to the generation of chimeric mouse models using homologous recombination (Capecchi, 1989).

Human ESCs (hESCs) were first derived and maintained in vitro in 1998 (Thomson et al., 1998), using 4-5 days old preimplantation embryos. Those cells possess certain defining characteristics that set them apart from somatic cells: instrinsic transcriptional hierarchy allowing the maintenance of pluripotency during self-renewal, a defining epigenetic state that keeps chromatin available for immediate cell fate decisions, and a cell cycle characterised by an extremely short G1 phase and the nearly absent checkpoint controls (Boheler, 2009).

1.6.1.1. Pluripotency maintenance in ESC

Pluripotency is transcriptionally governed by three key molecular regulators: OCT4, NANOG and SOX2. These transcription factors share a substantional segment of their target genes, and form a highly regulated circuitry in which these three genes control the transcription of one another, in addition to regulating other genes involved in pluripotency maintenance (Boyer et al., 2005). These genes are capable of both positive and negative regulation of transcription, creating a highly complex interaction network. This network consists of nuclear factors critical for maintenance of the ESCs in undifferentiated state and that possess a role in differentiation, representing a certain “cellular module” of pluripotency (Wang et al., 2006). Additional transcription factors implicated in pluripotency maintenance are listed in Table 1.8.

Transcription Factor	Function	Genomic location Comment
Oct4	Core circuitry	6p21.31 Homeobox gene
Sox2	Core circuitry	3q26.3-q27
Nanog	Core circuitry	12p13.31 Homeobox gene
Tcf3	Wnt signaling to core circuitry	19p13.3
Stat3	Lif signaling to core circuitry	17q21.31
Smad1	BMP signaling to core circuitry	4q31
Smad2 **Smad3**	TGF-β/Activin/Nodal signaling	18q21.1 15q22.33
c-Myc	Proliferation	8q24.21 Oncogene
Esrrb	Steroid hormone receptor	14q24.3
Sall4	Embryonic regulator	20q13.2 Zinc-finger protein
Tbx3	Mediates LIF signaling	12q24.1 T-box gene
Zfx	Self-renewal	Xp21.3 Zinc-finger protein
Ronin	Metabolism	16q22.1
Klf4	LIF signaling	9q31 Zinc-finger protein
Prdm14	ESC identity	8q13.3 Zinc-finger protein

Table 1.8 Key transcriptional factors involved in pluripotency maintenance. Table adapted from (Young, 2011).

Key epigenetic mechanisms involved in the maintenance of ESCs include nuclear architecture, defined chromatin structure, chromatin dynamics and histone modifications (Meshorer and Misteli, 2006). Nuclei of ESCs are characterised predominantly by euchromatin, and once differentiation is started and progresses, ESCs accumulate transcriptionally inactive heterochromatin. Thus, regulators of chromatin structure play an important role in embryonic development and in pluripotency maintenance.

Trithorax group (TrxG) and Polycomb group (PcG) protein complexes represent key molecular factors that regulate chromatic structure and contribute to pluripotency maintenance. In murine ESC, PcG proteins which possess H3K27-specific trimethylase activity directly repress developmental regulators that would lead to differentiation if their expressin were not inhibited, and all those regulators contain trimethylated lysine 27 on histone H3 (Boyer et al., 2006). In hESC, target genes of PcG are activated during cell differentiation, and OCT4, SOX2, and NANOG regulate a significant portion of these genes (Lee et al., 2006b). Conversely, trxG proteins have H3K4 trimethylase activity, and trimethylation of H3K4 (H3K4me3) leads to active chromatin state permisive for gene expression (Schuettengruber et al., 2007). The pattern of these two marks colocalising in ESC in clusters of developmental genes, such as homeobox genes, was named “bivalent domains”. Based on the presence of these marks, genes can be divided into three clusters: expressed genes, genes poised for expression, or repressed genes (Mikkelsen et al., 2007). Bivalency predisposes genes for both activation and inactivation, and during commitment towards differentiation, genes important in unrelated lineages lose their expression-permisive mark (Pietersen and van Lohuizen, 2008).

1.6.1.2. Concerns regarding clinical application of hESC

HESC are capable of long-term proliferation in vitro (Amit et al., 2000), and escape senescence through telomerase expression. However, long-term culture conditions lead to profound genomic and epigenetic changes of hESC lines (Maitra et al., 2005). Moreover, hESCs normally require a feeder layer in order to maintain undifferentiated phenotype. Traditionally, hESCs are thereby cultured on a a layer of gamma-irradiated murine fibroblasts that provide necessary stimuli for the maintanence of undifferentiated state, yet convey the risk of contamination with murine pathogens, including retroviruses, which renders them unsuitable for application in humans. In addition, other concerns regarding ESC application include tissue integration of transplanted hESC derivates, immunogenicity and graft rejection, as cells gradually become immunologically incompatible as differentiation progresses (Drukker et al., 2002), and tumourigenicity of hESC, discussed in greater detail in the following segment.

1.6.1.3. Human embryonic stem cells and cancer

The intimate link between ESC and cancer is highlighted by the historical fact that the field of embryonic stem cells research was initiated by experiments with embryonic carcinoma. The unique properties of hESC – indefinite proliferation and self-renewal – contribute to the tumourigenicity of hESC, and also represent hallmarks of cancer cells. Other shared traits include high telomerase activity, expression of oncogenes and lack of contact inhibition (Ben-David and Benvenisty, 2011).

Teratocarcinoma and teratoma are tumours comprised of both plutipotent stem cells and differentiated tissues, and the initial comparison of ESC with embryonic carcinoma cell lines demonstrating many shared properties was considered a validation of ESC pluripotency (Knoepfler, 2009). The most concerning discovery for regenerative medicine is the fact that ESC, a physiological counterpart of embryonic carcinoma cells, cause tumours. It seems likely that the removal of ICM from the embryo and establishment of ESC in an artificial environment of cell culture endows the ICM cells with novel properties, some of which are tumourigenic to a lesser or greater degree.

The so-called “teratoma assay”, a transplantation of pluripotent stem cells into mice in which the formation of benign tumours (teratomas) is then observed, is considered a demonstration of pluripotency of both ESC and iPSC (Wesselschmidt, 2011). Moreover, hESC can form aggressive undifferentiated tumours in mice (Shih et al., 2007), suggesting that tumourigenicity of ESC is considerably influenced by the host tissue, and thus highly unpredictable. Many regulatory elements and pathways are shared by ESC and cancer cells (Sperger et al., 2003, Ben-Porath et al., 2008), and established oncogenes such as MYC are required for the generation of iPSCs. If levels of MYC are lowered in iPSC, pluripotency of those cells decreases (Knoepfler, 2009). It may be deducted that the greater pluripotency, the greater tumourigenic potential, and if a stem cell were completely deprived of its ability to cause tumours, it would no longer be a stem cell. If we consider cancer a type of tissue, and ESC cells capable of differentiation into any kind of tissue, then inevitably these cells must be able to cause cancer.

Genome-wide expression profiling of hESC and human germ cell tumours has demonstrated many similarities between hESC and the human embryonal carcinoma cell, and a certain degree of transcriptome similarity with seminoma, a testicular tumour derived from germ cells (Sperger et al., 2003). Further expression data analysis has revealed that histologically poorly differentiated tumours overexpress ESC genes, and repress PcG-regulated genes, and key ESC markers (Nanog, Oct4, Sox2 and c-Myc) are more commonly overexpressed in poorly differentiated tumours in comparison with well-differentiated tumours (Ben-Porath et al., 2008).

1.6.1.4. Human embryonic stem cells in current clinical practice

In spite of serious ethical, religious and social dilemmas associated with human embryonic stem cells (hESC), laboratory results demonstrating that those can been successfully differentiated into clinically relevant tissues in vitro, including cardiomyocytes and dopamine neurons (Murry and Keller, 2008), encouraged clinical application of tissues engineered from those cells. In 2009, a clinical trial using oligodendrocytes derived from hESCs in patients with spinal injury began in the USA, and preliminary results have shown no serious adverse effects nor significant neurological improvement (Watson and Yeung, 2011), yet the study is still ongoing. Another clinical study of hESC-derived retinal pigment epithelium in patients suffering from macular dystrophy and dry age-related macular degeneration has demonstrated no adverse effects and modest improvement in the vision of selected patients (Schwartz et al., 2012). Nonetheless, while the therapeutic potential of stem cells seems practically unlimited, it still remains to be translated into effective clinical practice.

1.6.2. Induced pluripotent stem cells

Ever since the first hESC cell lines were derived from a human blastocyct in a process that required desctruction of the embryo, significant ethical concerns were raised over hESC (Vogel and Holden, 2008), and other above-described doubts regarding their clinical application inspired the creation of customised, personalised stem cells for individual patients using somatic-cell nuclear transfer (SCNT) (Robinton and Daley, 2012). SCNT has been previously used for “Dolly the sheep” cloning (Campbell et al., 1996), and pluripotent stem cells generated using SCNT would possess a patient’s genome and be used for generating any required tissue due to their pluripotency. However, since this procedure technically creates a human clone, although proven feasible for generating “personalised” ESC (Noggle et al., 2011), requirement of human oocytes and other technical and ethical considerations make SCNT-derived ESC unlikely for therapeutic use.

Induced pluripotent stem cells (iPSC) represent an alternative source of personalised stem cells. In 2006, Takahashi and Yamanaka introduced pluripotency in differentiated murine fibroblasts through the ectopic coexpression of four transcription factors by viral transduction (Takahashi and Yamanaka, 2006). IPSC derived from human differentiated cells quickly followed (Yu et al., 2007, Takahashi et al., 2007). The four factors used for reprogramming towards pluripotency included OCT4, SOX2, KLF4 and c-MYC used by Yamanaka’s group, while Thomson’s group avoided using c-MYC and KLF4, known oncogenes, and applied Nanong and Lin28 instead.

In spite of apparently shared traits between hESC and iPSC, the reprogramming process leads to copy-number variation (CNV) in the early passages of iPSC, leading to genetic mosaicism and selective disadvantage, as CNV decreases with further passaging (Hussein et al., 2011). Additionally, iPSC contain protein-coding point mutations (Gore et al., 2011) and aberrant methylation (Lister et al., 2011),

1.6.3. Mesenchymal stem cells

Mesenchymal stem cells, or multipotent mesenchymal stromal cells (MSCs) are multipotent progenitor cells capable of both self-renewal and differentiation towards cells with terminal functions. MSCs represent prototypical adult stem cells (in contrast to embryonic stem cells), and are distributed in various body tissues (Williams and Hare, 2011).

MSCs were first isolated from guinea-pig bone marrow as fibroblast-like cells capable of forming colonies (Friedenstein et al., 1970), and later were shown as able to differentiate into mesoderm-derived tissues (Friedenstein et al., 1974), and later emerged as important regulators of bone marrow microenvironment and haematopoietic stem cells niche (Mendez-Ferrer et al., 2010, Zhang et al., 2003). Outside of bone marrow, MSCs were shown to be present in most if not all adult tissues, and in addition to bone marrow, were isolated from adipose tissue, synovial membrane, the bronchi, umbilical cord blood, and peripheral blood. While MSCs from various sources are indeed heterogeneous, these cells share certain surface markers and mesodermal differentiation potential.

In terms of their multilineage potential, MSCs were shown to differentiate into osteoblasts, chondrocytes, adipocytes (Pittenger et al., 1999), myogenic cells, astrocytes and cardiomyocytes (Williams and Hare, 2011).

Mesenchymal stem cells are critical for tissue repair and regeneration. MSCs are rapidly recruited into the site of injury to promote healing and regeneration by differentiation and paracrine signalling (Hocking and Gibran, 2010). Given that tumours can be described, *cum grano salis,* as "wounds that never heal" (Dvorak, 1986), it is conceivable that the dramatic changes surrounding malignant tissue formation would attract MSCs in a manner similar to tissue wounds. The involvement of MSCs in cancer formation is described in the following segment.

1.6.3.1. Mesenchymal stem cells in cancer

Mesenchymal stem cells are an intergral part of stem cell niches of various organs. In spite of their potential and capacity to differentiate into various tissues such as fibroblasts, osteoblasts, and adipocytes (da Silva Meirelles et al., 2006, Pittenger et al., 1999), it has been shown that MSC are not inherently tumourigenic (Ljujic et al., 2013), but can contribute to pro-tumourigenic tissue microenvironment (Stagg, 2008). MSC can induce a local immunosuppressive effect or stimulate proinflammatory environment, and thus indirectly promote tumour growth (Uccelli et al., 2008). Bone-marrow derived MSC are recruited to the site of tumour formation, where proangiogenic stimuli these cells produce lead to increased tumour vasculature formation (Roorda et al., 2009). Higher tumour incidence has been reported if various tumour cell lines were co-implanted with MSC (Djouad et al., 2006, Yu et al., 2008), and most importantly, a higher recurrence rate of haematologic malignancies has been noted among patients co-transplanted with MSC (Ning et al., 2008). Additionally, stromal MSC have been shown to promote breast cancer metastases (Karnoub et al., 2007), and differentiate into carcinoma-associated fibroblasts that further sustain tumour growth (Mishra et al., 2008).

However, there is evidence showing antitumour effects of MSC, suggesting their role in cancer formation and growth may be context-dependent: in Kaposi sarcoma model, MSC were recruited to sites of tumour formation as expected, but inhibited tumour growth through Akt inactivation (Khakoo et al., 2006). A study in hepatoma and breast carcinoma model has shown similar effects due to Wnt pathway inactivation by MSC (Qiao et al., 2008a, Qiao et al., 2008b).

These data suggest that the ultimate effect of MSC on tumour biology may be highly dependent on individual molecular properties of given tumour cells, systemic and local immunological response and further factors that remain to be unravelled.

1.6.4. Human epidermal stem cells

Epidermal stem cells (EpSC) are located in epidermal niches, and their primary role in skin homeostasis is the replenishment of lost cutaneous cells (Shen et al., 2013). While there are extensive data describing the biology of EpSC in mouse skin, very little is known about human EpSC.

While keratin 15 is routinely used to isolate epidermal stem cells from murine hair follicles, it is not equivalently efficient marker for human EpSC (Ohyama, 2007), for which other markers have been proposed, including CD200 (Ohyama et al., 2006, Garza et al., 2011).

Interfollicular epidermis that predominates in humans compared to mice may contain EpSC in so-called rete ridges.

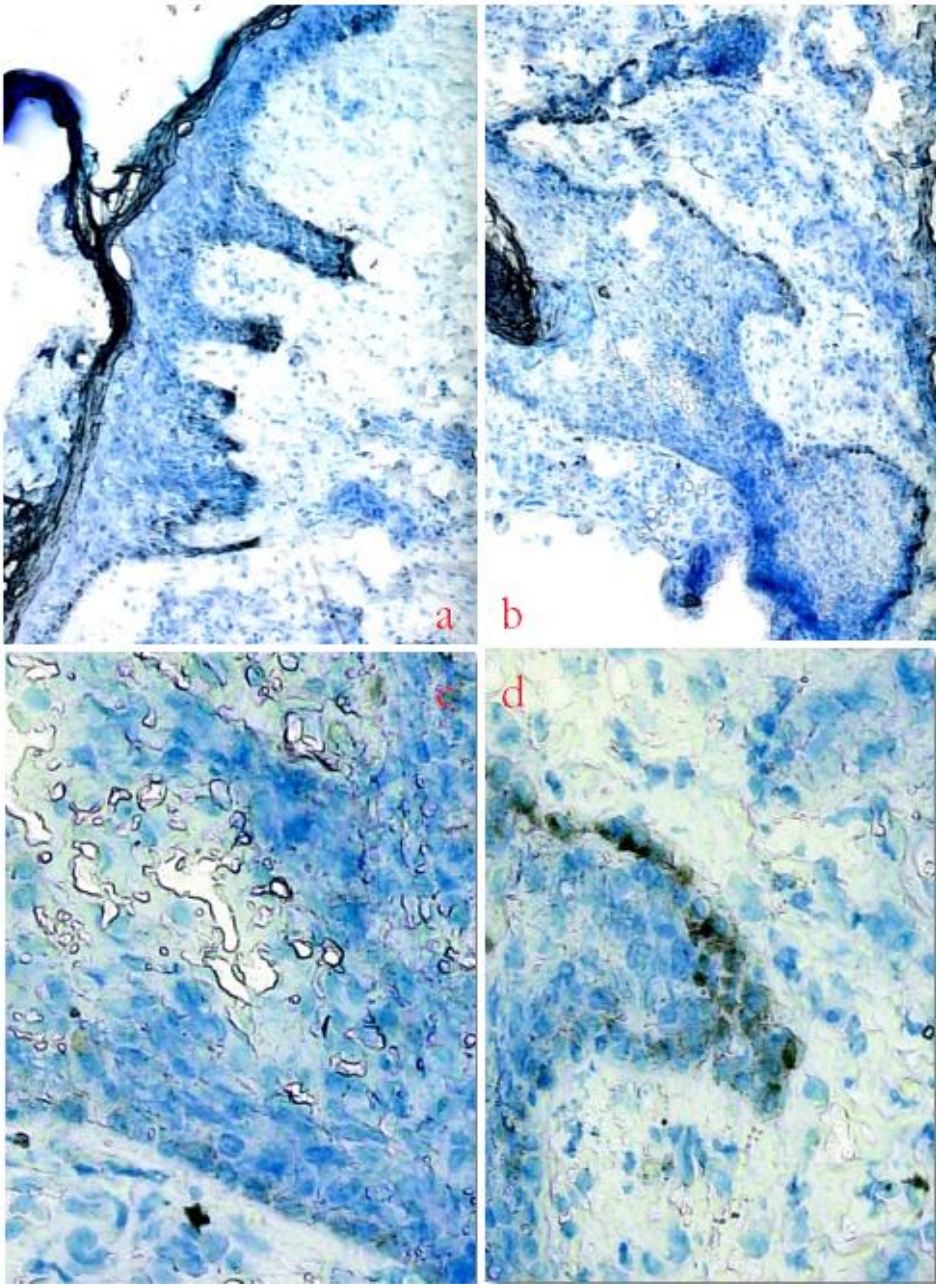


Figure 1.23 Panel of skin samples stained with methylene blue depicting rete ridges. a) 5x magnification b) 10x magnification c) + d) 40x magnification. Suggested EpSC are present at the demarcating edge of rete ridges and stain very darkly.

These cells were suggested to express high levels of beta1-integrin, melanoma chondroitin sulfate proteoglycan and Lrig1 (Boehnke et al., 2012). Alternative markers proposed included α6 integrin and low expression of CD71 (Webb et al., 2004). This methodological discrepancy may be due to the conceptual definition of stem cells, and technological limitations of antibody staining. Currently, there is no definite consensus with respect to the location, identifying markers and biological characteristics of human interfollicular EpSC.

Cancer stem cells

One of the central questions in cancer biology is the cell of origin of cancer, which to date has not been answered. Given that each subtype of cancer is markedly different in terms of intrinsic histological features, propensity to metastasise and response to therapy, it may be conceived that such variety reflects different cells of origin for these tumours, each driving the distinct phenotype ultimately harboured by cells of the tumour (Visvader, 2011). Such cell of origin is expected to be a normal cell that acquires tumourigenic mutations which form a hierarchically organised tumour mass by clonal evolution, and this mass is then maintained by a distinct subpopulation of cells capable of both self-renewal and generating other cells of the neoplasia. This concept presumes the existence of two distinct populations within malignancies: cell of origin (tumour-initiating cell(s)), which leads to tumour formation and cancer stem cells (CSC), responsible for subsequent tumour sustenance. This concept is depicted in Figure 1.15.

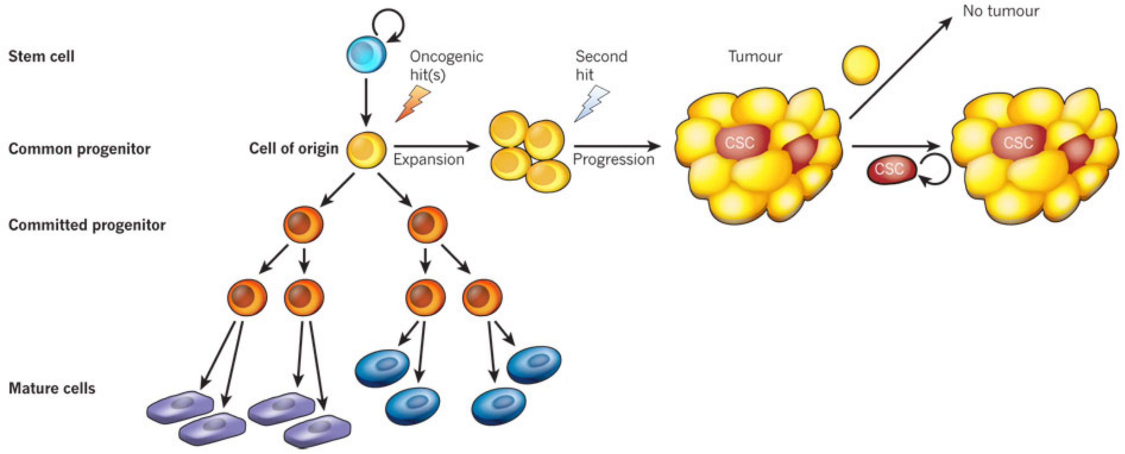


Figure 1.24 Hierarchical structure of tissue cells linked with neoplastic initiation and progression. According to this model, only the CSCs sustain tumourigenesis, and are distinct from the cell of origin. Figure taken from (Visvader, 2011).

The theory of a single cell of origin of cancer has been postulated in 1976: acquired genetic variability of the original tumourigenic cell allows for clonal selection towards more aggressive sublines, and this evolution-like selection process leads to the genetic and biological variety within tumours (Nowell, 1976). This parallel with Darwinian selection process suggests that cancer is a quasi-species endowed with adaptive properties (Greaves and Maley, 2012). According to this theory, tissue is the environmental context for cancer evolution, and various genetic subclones within the tumour represent adaptive solutions to (micro)environmental constraints: hypoxia, space limitations, immune system, nutrition, and most importantly cancer therapeutics. The genotoxic character of many therapeutic agents provides selective pressure for the growth of clones that gain further mutations accounting for treatment resistance and increased aggressiveness of many recurrent cancers. Presuming that mutagenic processes are random and non-purposeful, clonal evolution oversees the interplay between "driver" lesions that drive clonal expansion and confer a selective advantage, "passenger" or "hitchhiker" lesions which remain selectively neutral, and “mutator” lesions that lead to other genetic changes in the context of tissue microenvironment. Importantly, epigenetic changes which are much more readily inducible compared to genomic mutations may be key drivers of clonal evolution (Siegmund et al., 2009) and tumour heterogeneity (Varley et al., 2009). Single-cell sequencing in breast cancer has shown clonal evolution of the tumour based on genetic genealogy of dominant clones and rare, intermediate clones that seem to have been overcome in competition with ultimately prevalent clones (Navin et al., 2011), and data from leukemia further confirmed evolutionary pattern of cancer development and its clonal architecture (Anderson et al., 2011, Campbell et al., 2008, Ding et al., 2012). The clonal and evolutionary character of metastases was described in pancreatic carcinoma, demonstrating that metastatic clones are the progeny of non-metastatic parental clones within the primary tumour mass (Yachida et al., 2010), and in renal cell carcinoma (Gerlinger et al., 2012). Clonal evolution of cancer through selective pressure is shown in Figure 1.16.

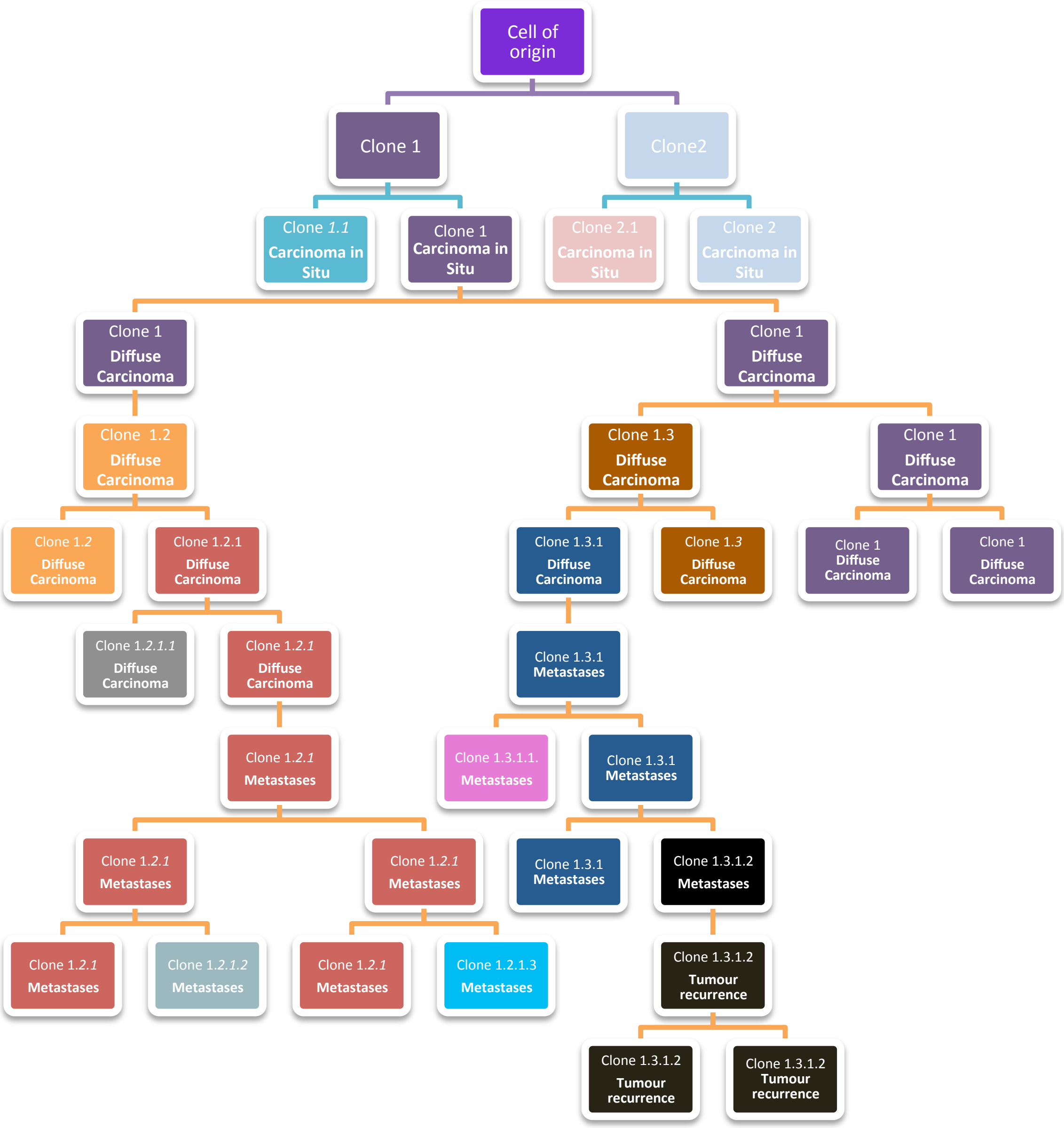


Figure 1.25 Clonal evolution of cancer. Change of colour indicates mutation. Figure adapted from (Greaves and Maley, 2012).

In consistence with this model, cancer stem cells would represent a subclone with distinct evolutionary advantage that allows CSC to supercede "ecological pressure" of radiation and chemotherapy and eventually reform the tumour mass. However, cancer stem cells would not be necessarily equal to the cell of origin of cancer, although this feature is often attributed to CSC[4].

The first CSC were identified in acute myeloid leukemia as a rare cell subpopulation definable by surface markers and capable of inducing tumour in nonobese diabetic (NOD)-severe combined

---

[4] This may be perhaps due to mostly semantic reasons rather than biological evidence.

immunodeficient (SCID) mice (Lapidot et al., 1994). This approach became a paradigm for CSC studies, in which putative CSCs are isolated from ‘bulk’ cancer cells using fluorescence-activated cell sorting (FACS) using antibodies against defined cellular surface markers, and subsequently transplanted into immunocompromised mice (Rosen and Jordan, 2009). However, this experimental approach does not take into account tumour microenvironment and the immune system as critical contributors to tumourigenesis capable of conferring phenotypic and functional differences to cancer cells (Polyak et al., 2009). Additionally, if cancer is a constantly evolving quasi-species, it seems very likely that CSC evolve along each tumour in consistence with selective pressure such as cancer therapy (Reya et al., 2001). Distinct pools of CSCs were identified in various sections of glioblastoma, yet demonstrating a shared ancestor (Piccirillo et al., 2009), further corroborating the concept of CSC as a subclone within a clonal hierarchy.

In spite of widespread debates and doubts about the sole existence of CSC (Rosen and Jordan, 2009), these have been reported in various tumours using a variety of surface markers (Magee et al., 2012, Karamboulas and Ailles, 2013), summarised in Table 1.5.

Tumour	Positive surface marker	Negative surface marker
Breast cancer	CD44	CD24
Colon cancer	CD44, EpCAM CD133	
Glioblastoma	CD133	
Head and neck SCC	CD44	
Leukemia	CD34	CD38
Liver cancer	CD133, CD44 CD133, CD24	
Lung cancer	CD133	
Melanoma	ABCB5	
Ovarian cancer	CD44, MyD88 CD133	
Pancreatic cancer	CD44, CD24, EpCAM	
Prostate cancer	CD44, α2β1, CD133	

Table 1.9 Various tumours in which CSC have been identified using defined surface markers.

In cutaneous SCC, CD133 has been proposed as a marker for CSC based on short term primary cSCC cultures transplanted into mice, which represents a novel technological approach (Patel et al., 2012). However, this marker was not validated in a subsequent study, which proposed beta-1 integrin as a CSC marker in cSCC instead (Dallaglio et al., 2013). It is important to note that both studies used primary cSCC rather than stable cell lines.

While the concept of cancer stem cells has received considerable attention during the past decade, its direct role in therapeutic application and patient management has yet to be established.

1.7. Analytical approaches to large scale genomic data

While collection of data used to represent major hindrance in scientific progress in the past, emerging novel tools in genomic and epigenomic exploration enable generation of large amounts of data within a relatively short time; however, such data often become serious analytical challenges. Novel tools are constantly being developed to enable more precise and less computer-skills demanding analysis of large-scale genomic data.

1.7.1. Microarrays

Since their first introduction over fifteen years ago, microarrays have by now become a practically indispensible tool in many research areas (Brown and Botstein, 1999, Diehn et al., 2000), including gene expression, DNA copy number variation, DNA methylation profiling, and more recently, microRNA profiling. Microarrays are usually comprised of a glass slide with fragments of nucleic acids mounted on to precise locations on the slide. Upon DNA or RNA hybridisation, corresponding short molecules coupled with a light-emitting element bind to their corresponding counterparts in the hybridised material. Eventually, those features form spots that are read by a special reader and processed as an image in a computer.

Due to complexity and data scale of microarray experiments, standard statistical approaches and tools are not be adequate for accurate data processing (Zhang, 1999). Instead, microarray data processing depends on a number of computational steps, including data extraction, data storage, quality control, data normalisation and feature annotation. The analytical precision depends on many factors, such as background noise and quality of the input material.

Many internal and external factors need to be taken into account in microarray data analysis: given that data are acquired by image processing and spot quantification after laser illumination, the process of data acquisition is thereby a major source of variability, since the chemistry of the fluorescent labels depends on many stochastic factors (including the humidity and temperature of the surrounding environment). Hybridisation is also dependent on various random elements such as pipetting precision, and the task of proper analysis is to remove all those confounding factors to the maximum extent possible. Both pre- and post-hybridisation steps can lead to substantial microarray data variability (Ahmed et al., 2004).

1.7.2. Image analysis and quality control of microarray data

Image analysis is the first step of microarray data processing that fundamentally determines all subsequent downstream steps (Kadanga et al., 2008). Key steps in overcoming issues of background noise, spot shape, size and position include spot detection, signal to background segmentation, and detecting signal intensity and signal quality (Wang et al., 2001).

Following this first task in microarray data processing, quality control removes spots or whole arrays from the set on which the nucleic acid has not hybridised properly, or arrays that bear a manufacturing defect. This is usually conducted through an array of visual assessments of the data, or through quality scores generated for individual array spots. Only those spots and arrays that pass a rigorous quality control should undergo the next step of normalisation.

1.7.3. Normalisation

This process removes experimental sources of variability (systemic bias). It is an essential pre-processing procedure that can remove noise, systematic variation due to different characteristics of fluorescent probes, slide scanner efficiency or labelling methods (Bilban et al., 2002). If the probes intensity is properly normalised, systemic variation is removed and the data become more robust. Also, normalisation allows comparison of data derived from different experiments.

Various normalisation methods exist that take into account individual spot size, signal-to-noise ratios, background noise and saturation status (Kadanga et al., 2008). A *global method* of adjustment is applied to the whole dataset, such as quantile normalisation (Gagnon-Bartsch and Speed, 2012). *Application-specific* methods, on the other take into account details of a given experiment, such as batch number.

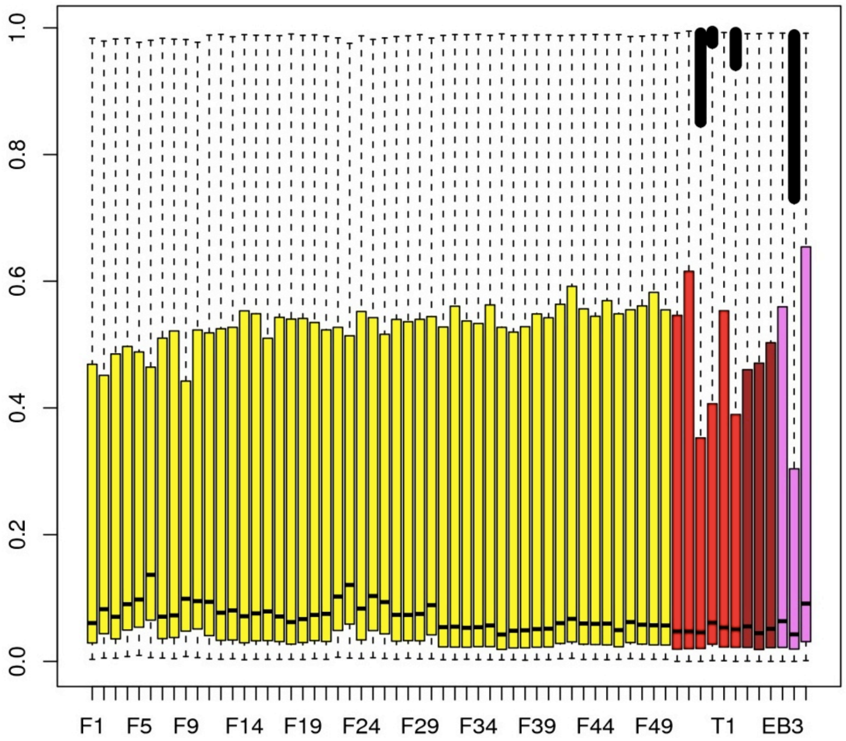


Figure 1.26 Boxplots of methylation data pre-normalisation. Y axis=beta values (methylation level). Yellow boxes=normal skin, red boxes=cSCC cell lines (non-RDEB), pink boxes=cSCC cell lines (RDEB), brown boxes=PHK. Boxplots show uneven distribution of beta-values across samples.

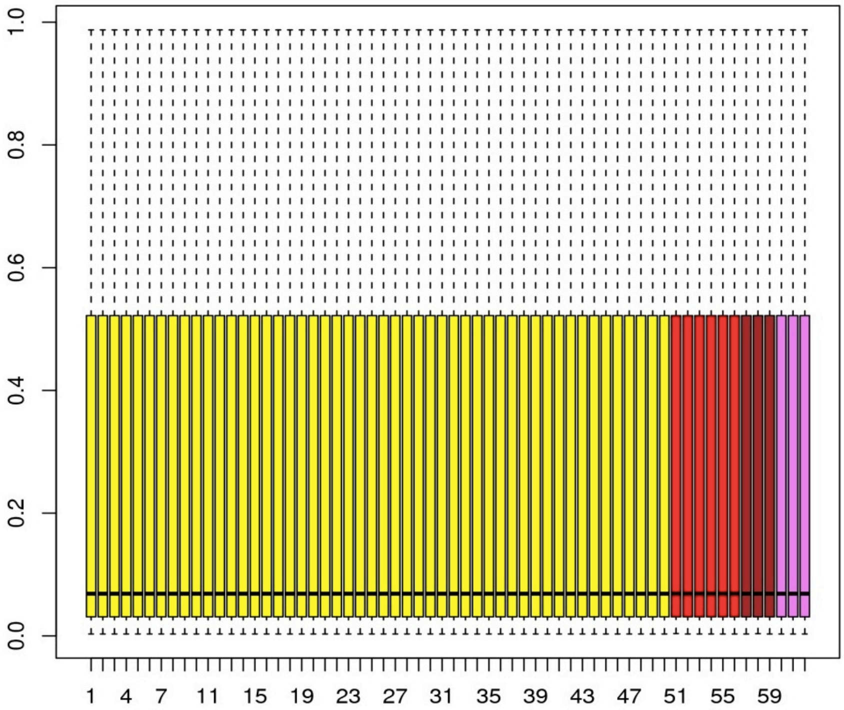


Figure 1.27 Boxplots of methylation data post-quantile normalisation. Y axis=beta values (methylation level). Yellow boxes=normal skin, red boxes=cSCC cell lines (non-RDEB), pink boxes=cSCC cell lines (RDEB), brown boxes=PHK. All boxplots now display even beta-values.

1.7.4. Downstream analysis

Once the data has been normalised, the process of identifying differentially expressed or methylated genes involves non-parametric (Bayesian) statistics, multiple comparisons and adjustment of p-values for given false discovery rate (usually 0.05).

Genes and samples can be clustered into groups based on the mathematically defined similarity between the individual values. The process of unsupervised hierarchical clustering can be very useful in finding groups of tumours sharing common features, on the other hand, those results can also be the result of an artefact, and must always be approached critically.

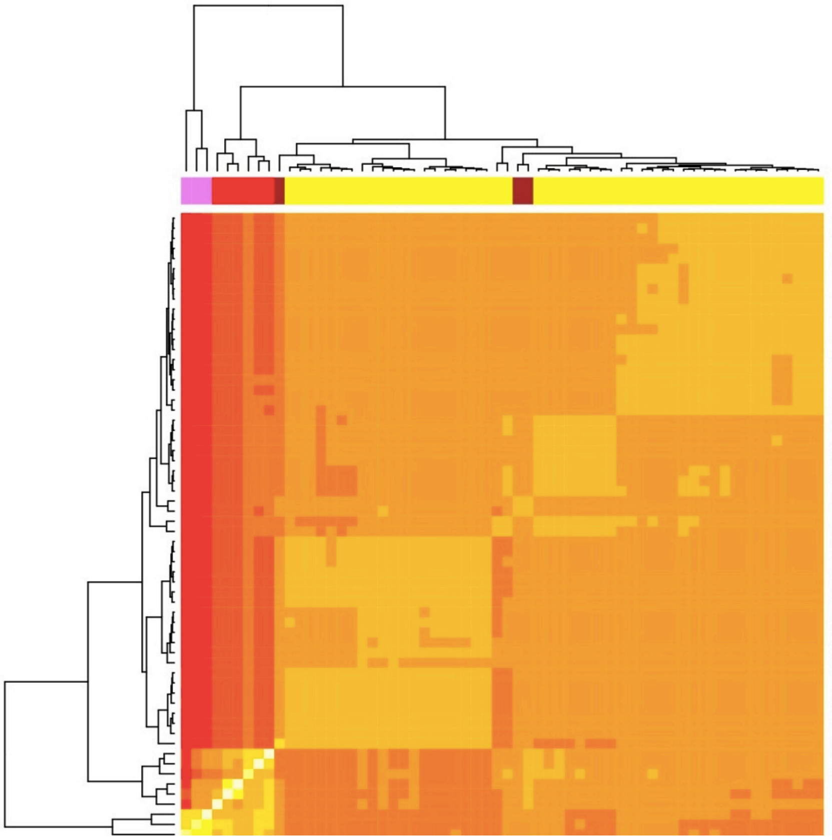


Figure 1.28 Horizontal and vertical clustering of normalised methylation data based on Pearson's correlation. Dendrograms display relative distance between the correlation values. Yellow boxes=normal skin, red boxes=cSCC cell lines (non-RDEB), pink boxes=cSCC cell lines (RDEB), brown boxes=PHK. Primary normal human keratinocytes (PHK) cluster with normal skin samples.

1.7.5. Bioconductor

In the recent year, a specialised tool for the analysis of microarray data was developed by the efforts of biostatisticians concentrated around the statistical language R. The project Bioconductor, a statistical tool for processing high-throughoutput genomic data is an open-source, command-line driven programme that uses several hundred of packages (coded programs) for processing raw input data (Gentleman et al., 2004). Currently, Bioconductor is capable of processing SNP, expression, methylation microarray data and sequencing data with dedicated packages using standardised statistical approaches.

The main advantage of open-source software is its availability for anyone with internet access, complete user control over each step of analytical pipeline, and also transparency and reproducibility of results at every stage of data proccesing in comparison with "black box" proprietary software. The main disadvantage is a high demand on user's computational and statistics knowledge, and the process of data analysis can often be rather time-consuming. In spite of these inherent drawbacks, Bioconductor provides robust, transparent, comprehensive and up-to-date tools for data analysis that are especially useful for analysing microarray data.

1.8. Hypothesis and Aims

Dysregulation of diverse epigenetic processes is a hallmark of malignant onset and progression in many tumours. A significant proportion of cSCC arises from precursor lesions (AK), and this process may be at least partially mediated by a variety of epigenetic changes. In order to test this hypothesis, I aimed at answering the following questions:

1. Is dysregulated methylation an integral component of the pathological progression from normal skin to cSCC, and if so, is this reflected in changes in the tissue expression profiles?
2. Are there fundamental differences in methylation and expression profiles between standard cultured keratinocytes and established cSCC cell lines, and if so, how do these compare to clinical samples?
3. Is there a transdifferentiation or dedifferentiation pattern in cSCC methylation compared to normal keratinocytes that allows cancer cells to obtain novel cancer-specific properties?
4. Which miRNAs are differentially expressed between non-tumour skin adjacent to cSCC and clinical samples of cSCC, and are these responsible for changes in the expression profile?

To answer these questions, I carried out the following experiments:

i. A genome-wide methylation examination of normal skin, AK and cSCC tumour biopsies accompanied by integration with genome-wide expression profiling after stratification for age, immune status and gender. Key differences in methylation between the normal skin and cSCC were validated using a whole-genome bisulfite sequencing.
ii. A genome-wide methylation profiling of cSCC cell lines, cultured keratinocytes, accompanied by genome-wide expression profiling and integration of the obtained data.
iii. A comparative analysis of methylation pattern of cSCC cell lines and cultured keratinocyte with a variety of differentiated and undifferentiated tissues including ESC.
iv. MiRNA microarray profiling of cSCC samples and matched adjacent normal skin coupled with microarray expression data.

This approach will provide novel insights into epigenetic regulation of critical cellular processes in this skin malignancy and may identify novel targets directly involved in their onset and progression.

Human papillomavirus (HPV) is commonly present in non-melanoma skin cancer, and I hypothesised that HPV may cause pro-oncogenic changes in methylation in cSCC samples. Additionally, I hypothesised that a new DNA polyoma virus (MCPyV) discovered in a different type of non-melanoma skin cancer (Merkel cell carcinoma) may be present in higher proportion in non-melanoma skin cancer samples of organ transplant recipients, and lead to epigenetic changes in cSCC samples. This hypothesis leads to the following questions:

5. Is HPV presence in cSCC associated with unique changes in methylation?
6. Is MCPyV more prevalent in non-melanoma skin samples of transplant recipients?
7. Are there unique methylation changes in cSCC samples that harbour the virus?

To answer these questions, I carried out the following experiments:

v. Genome-wide methylation profiling of HPV-positive and negative cSCC samples.
vi. MCPyV profiling of various non-melanoma skin lesions derived from both transplant and non-transplant patients.
vii. Genome-wide methylation profiling of MCPyV-positive and negative cSCC samples.

This approach may identify distinct molecular changes in MCPyV-positive tumours with potential implications for prevention or therapeutic options.

2. General Materials and Methods

This chapter describes materials and methods that are common to the majority of the work presented in this thesis. Methods that pertain uniquely to specific chapters are outlined within these individual chapters.

2.1. Tumour Samples

Tumour samples were obtained from consented patients between 1995-2013 in Dermatology clinics at Barts and the London NHS Trust (London, United Kingdom). Prior to collection, all patients provided informed written consent in accordance with ethical approval from the East London and City Health Authority local ethics committee.

During the procedure, patients received a local anaesthetic to the tumour area, and 4-6mm punch biopsy was excised from the tumour mass and immediately snap-frozen in liquid nitrogen. Samples were subsequently stored in -80°C. Adjacent normal skin was collected during subsequent tumour excision procedure in the operating theatre, and was also immediately snap-frozen in liquid nitrogen and stored in -80°C.

Paraffin-embedded tumour samples were obtained from the Department of Pathology, Barts and the London NHS Trust (London, United Kingdom).

2.1.1. Histological Staining of Tumour Sections

2.1.1.1. Haematoxylin & Eosin staining

Tumour biopsies were embedded in OCT (optimal cutting temperature compound, VWR, UK), mounted on cork discs and labelled. Samples were cross-sectioned with an OTF5000 cryostat (Bright Instrument Company Ltd., UK) into 10 µm thick slices and immediately mounted on Superfrost plus slides (VWR, United Kingdom) and stored in -80°C prior to staining. Sections were equilibrated to ambient temperature for approximately 2 minutes, continuously rehydrated in 100%, 90% and 70% ethanol for 2 minutes each and washed 5x in distilled water. Slides were then submerged in haematoxylin for 4 minutes in Mayer's haematoxylin solution (Sigma-Aldrich, United Kingdom) to stain the basophilic elements of cells. Slides were then submerged in running tap water for 8-10 minutes, washed once in 1% hydrochloric acid solution for differentiation and counterstained with eosin Y (Sigma-Aldrich, United Kingdom) for 1 minute. Slides were then washed in distilled water and dehydrated by a series of ethanol of increasing concentration (70%, 90% 100%) and fixed in xylene. Sections were mounted in DPX mountant (Sigma-Aldrich, United Kingdom) and covered with coverslips.

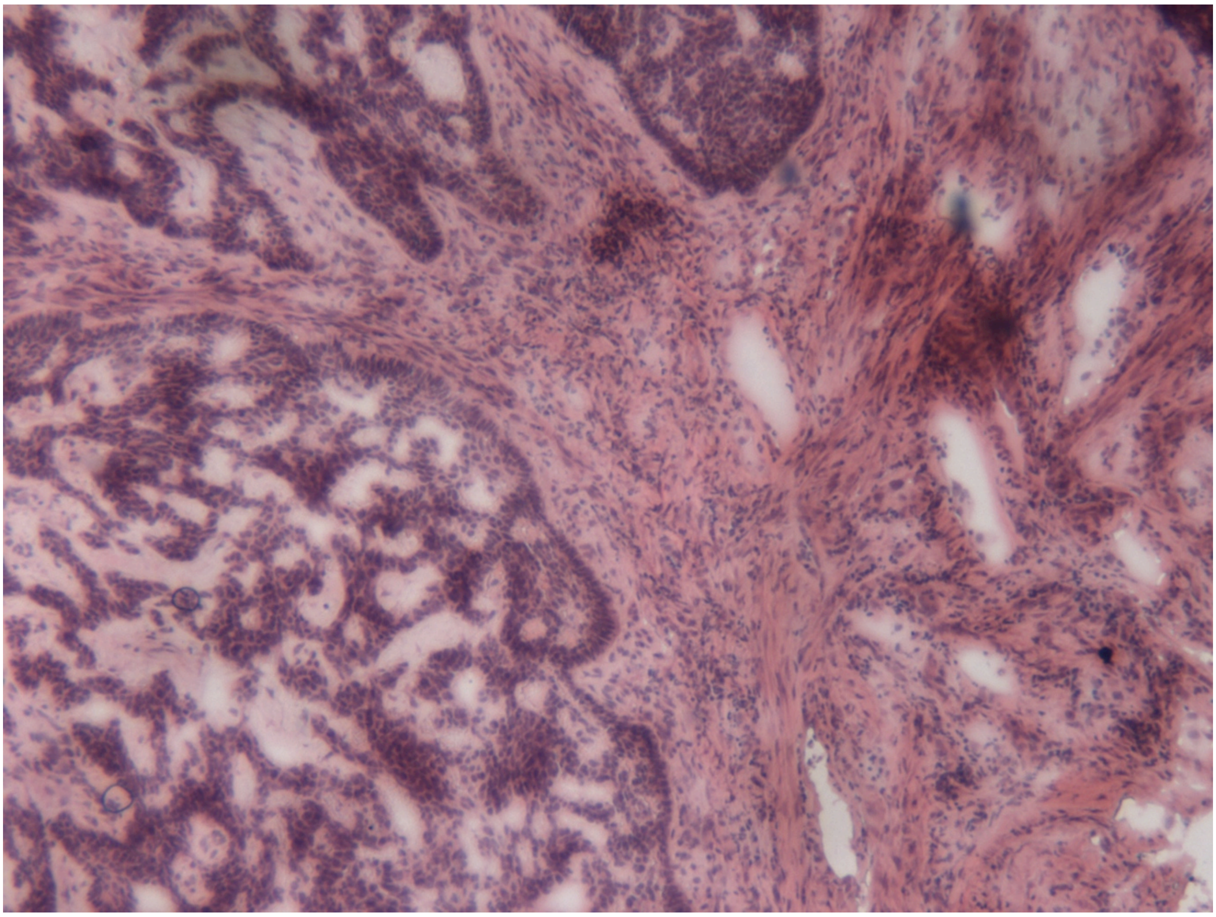


Figure 2.1 H&E stained section of a cSCC sample.

2.1.1.2. Acid fuchsine and Toluidine blue staining

Acid fuchsine solution was prepared by dissolving 0.25 g of acid fuchsine powder (Sigma-Aldrich, United Kingdom) in 50.0 ml of distilled water. Toluidine blue dye solution was prepared by dissolving 0.5 g in 50.0 ml of distilled water.

Sections mounted to Superfrost plus slides (VWR, United Kingdom) and stored at -80°C prior to staining were equilibrated to ambient temperature for approximately 2 minutes, continuously rehydrated in 100%, 90% and 70% ethanol for 2 minutes each and washed 5x in distilled water. Slides were covered with toluidine blue solution for 2 minutes to stain basophilic structures, washed in distilled water and covered with acid fuchsine solution for additional 2 minutes to counterstain eosinophilic elements in the cells. Sections were then washed in distilled water to remove excessive stain, air-dried and mounted in DPX mountant (Sigma-Aldrich, United Kingdom) and covered with coverslips.

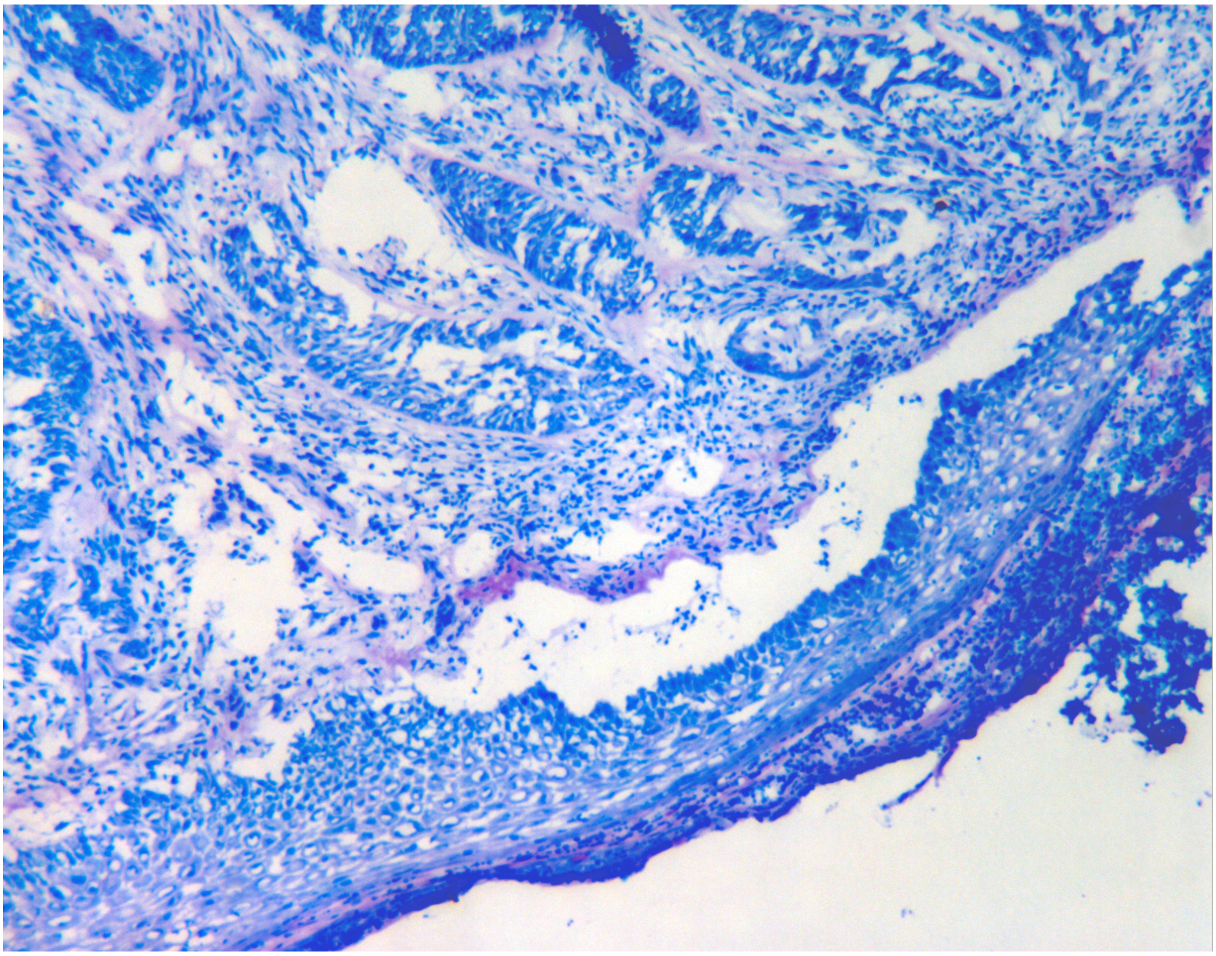


Figure 2.2 Acid fuchsine and toluidine blue-stained section of a cSCC sample.

2.2. DNA-based techniques

2.2.1. Agarose gel electrophoresis

The integrity of DNA isolated from fresh-frozen tissue (see below for details) and products of PCR reactions were visualised on 1% agarose gel containing ethidium bromide (6 µl of ethidium bromide in 100 ml of 1% agarose gel solution) in Syngene G:BOX F3. The gel was prepared by mixing 1g of agarose powder (Sigma Aldrich, United Kingdom) and 1% TBE buffer, after brief swirling the mixture was microwaved until the agarose powder was fully dissolved. 6 µl of ethidium bromide were added to the mixture before pouring it to the gel tanks and inserting well combs. Gels were allowed to cool for 1 hour. Ten microliters of PCR products were mixed with 3 microliters of loading dye and carefully pipetted into gel wells. Pipette tip was changed between samples. Six microliters of DNA ladder mixed with loading dye were added into the last well. Samples were run in 1x TBE buffer for 1 hour at 100 V (unless otherwise specified). Images of products on gels were acquired with GeneSys software.

2.2.2. Laser capture microdissection (LCM) of frozen tissue sections with subsequent DNA isolation

Tumour biopsies and adjacent normal skin were embedded in OCT (optimal cutting temperature compound, VWR, UK), mounted to cork discs and properly labelled. Samples were cross-sectioned with OTF5000 cryostat (Bright Instrument Company Ltd., UK) into 10 µm thick slices and immediately mounted on pre-chilled 1.0 mm polyethylene napthalate (PEN) coated membrane slides (Carl Zeiss Ltd., UK). PEN membrane slides had been pre-treated with 1500 mJ of UVB light for 40 minutes in order to enhance adhesion of tissue to the membrane.

Immediately prior to LCM, samples were fixed with 70% alcohol for 3 minutes, washed in distilled water for five times, and stained with 0.5% acid fuchsine solution for 1 minute and counterstained with 0.5% toluidine blue solution for another minute to visualise microscopic features. Samples were immediately mounted in PALM Microbeam laser capture microscope (Carl Zeiss Ltd., Germany) slide holder. 10x magnification was used to detect target cells. Those cells were then hand-labelled on computer using a mouse cursor, and excised by the machine's UV laser. The excised cells were collected with a needle into sterile Eppendorf tubes containing 20 µl Proteinase K and 180 ATL tissue lysis buffer (Qiagen, UK), incubated for 24-48 hours at 56°C. DNA was extracted using the standard protocol QIAamp DNA micro kit (Qiagen, UK), and eluted in 50 µl of sterile water. DNA integrity was assessed by agarose gel electrophoresis: 1 µl of DNA was loaded into wells of agarose gel, run for 30 minutes at 100 volts (V), and visualised using UV light illumination (Figure 2.3).

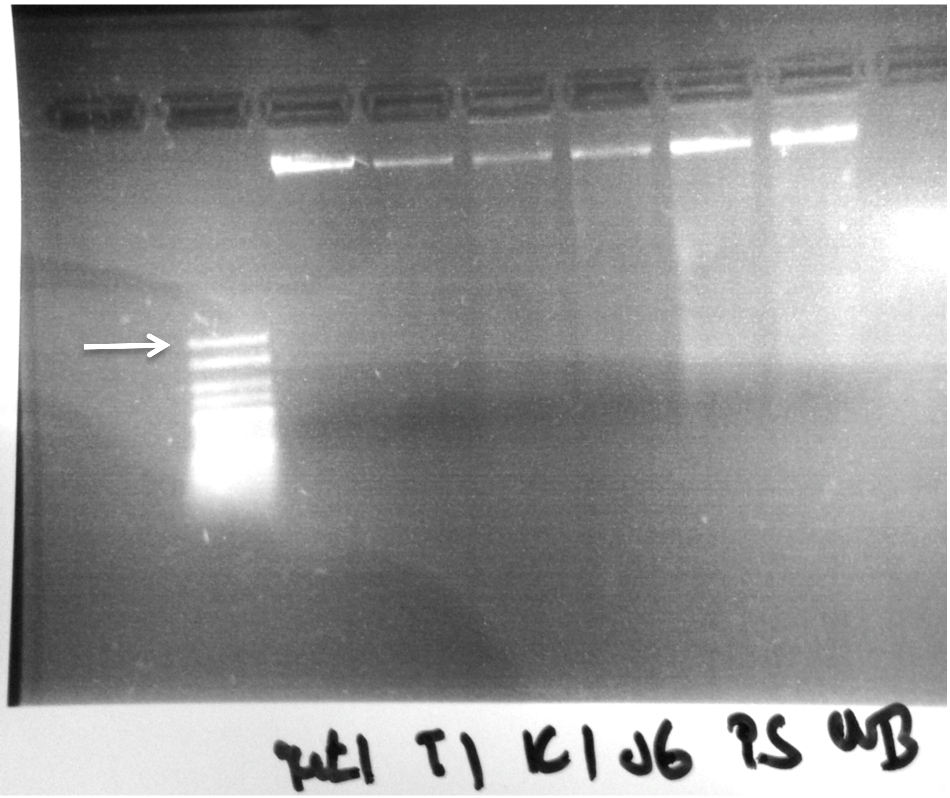


Figure 2.3 DNA integrity assessment by agarose gel electrophoresis. Agarose gel electrophoresis of DNA extracted from cultured cSCC cell lines (MET1, T1, IC1, JG, PS, WB, lanes 3-8 from left) and 100 bp ladder in lane 2 (white arrow indicates 1000 bp). This figure demonstrates intact DNA integrity.

2.2.3. DNA isolation from formalin fixed paraffin embedded tissue

Ten sections of 10 μm thickness were cut onto individual SuperFrost slides (VWR, UK) using a microtome. Sections were then deparaffinised using xylene for 2x for 5 minutes, dipped in distilled water and dehydrated with 100% ethanol. Samples were then rehydrated with 70% ethanol and scraped into 1.5 ml Eppendorf tube using a sterile blade. 20 μl Proteinase K and 180 ATL tissue lysis buffer (Qiagen, UK) were added to the tubes, and those were then vortexed 4x for 10 seconds and incubated for 24-48 hours at 56°C. DNA was extracted using the standard protocol QIAamp DNA micro kit (Qiagen, UK), and eluted in 100 μl of sterile water.

DNA isolated from FFPE tissues is often fragmented (Figure 2.4). To confirm DNA integrity, DNA polymerase chain reaction (PCR) was performed using the following set of primers for the β-globin gene: B1 (5'-GTGTGCTGGCCCATCACTTT-3') and B19 (5'-CAAGAAAGCGAGCTTAGTGA-3'). The reaction yields an amplicon of 120 bp in length (Hiyama et al., 1990), which was assessed using 1.5% agarose gel electrophoresis described below. Only those samples that produced the target amplicon during this initial step were used in subsequent downstream analysis.

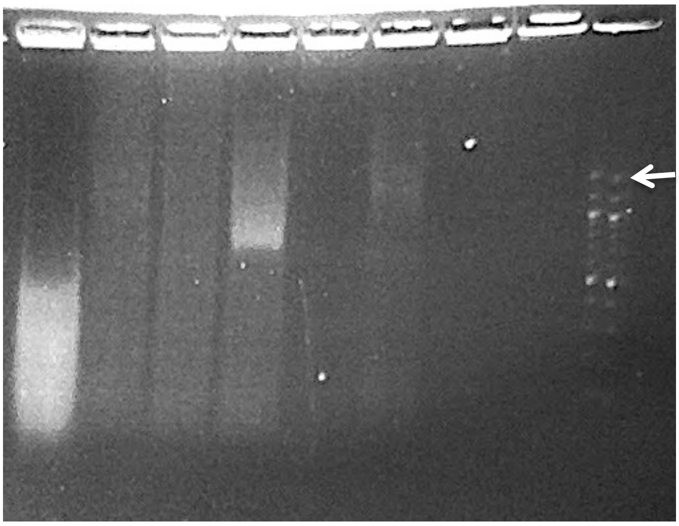


Figure 2.4 Agarose gel electrophoresis of DNA isolated from FFPE samples. From left: Lane 1-6 contain DNA isolated from FFPE non-melanoma skin cancer samples, lane 9 contains 100 bp ladder (white arrow corresponds to 1000 bp). This figure shows high fragmentation of DNA isolated from FFPE samples.

2.2.4. DNA isolation from cultured cells

2.2.4.1. DNA isolation from primary cells and established cell lines

Both primary cells and established cell lines were grown to 70%-90% confluency in tissue culture flasks. The culture media was removed and replaced with phosphate buffered saline (PBS) for 5 minutes in order to remove remaining media. The cells were then covered with 0.05% trypsin and 0.02% EDTA in PBS and incubated at 37°C for 5-20 minutes. Once detachment of cells was confirmed under optical inverted microscope, 5-15 ml of culture media were added to the flasks and the content was transferred into 15 or 50 ml centrifuge tube and spun at 2500 rpm for 5 minutes. The supernatant was then removed, and the cell pellet was immediately frozen on dry ice and stored in -80°C. DNA was extracted from the cells using QIAamp DNA mini kit (Qiagen, UK) following the standard protocol and eluted in 200 µl of sterile water.

2.2.4.2. DNA isolation from fluorescence-activated cell sorting (FACS) isolated cells

FACS-sorted cells were allowed to attach in 6-well plates for 12 hours (see below details on FACS sorted cells isolation and cell culture). Cells were then washed twice with PBS, incubated with 1.5 ml of Accutase® solution (Sigma, United Kingdom) at 37°C until detached from the well bottom. 1.5 ml of culture media was added, and the well content was collected into 15 ml centrifuge tube. Cells were pelleted by 5 minutes of centrifugation at 1200 rpm, the supernatant was removed and cells were immediately frozen on dry ice. Prior to use, DNA was extracted in parallel with RNA using the AllPrep DNA/RNA Micro Kit (Qiagen, UK) following a standard protocol described in greater detail in Chapter 2.4.2.2. DNA was eluted from the column into 50 µl of water.

2.2.5. DNA isolation from tissue samples

DNA was isolated from fresh frozen tissues stored at -80°C. Between 1g and 5g of tissue was cut into small pieces on dry ice, transferred into 1.5 ml pre-chilled Eppendorf tube and covered with 180 µl ATL lysis buffer and 20 µl Proteinase K (Qiagen, United Kingdom) and repeatedly vortexed. Samples were incubated at 56°C for 24-72 hours, with additional 20 µl Proteinase K added after 48 hours of incubation. DNA was extracted using the standard protocol QIAamp DNA micro kit (Qiagen, UK), and eluted in 100 µl of sterile water.

2.2.5.1. Archival DNA isolated from skin lesions

Professor Catherine Harwood, Drs Karin Purdie and Victoria Brown have isolated DNA from various skin specimen for the purpose of previous projects using respective total DNA isolation protocols (Harwood et al., 2000, Brown et al., 2004). The DNA was stored in -80°C for 3-12 years.

For the purpose of subsequent DNA polymerase chain reaction (PCR) experiments, DNA integrity of the archival nucleic acid isolated from a tissue sample was confirmed using the following set of primers: KM29 (5'-GGTTGGCCAATCTACTCCCAGG-3') and GH21 (5'-GGAAAATAGACCAATAGGCAG-3'), which yield an amplicon of 345 bp in length. The PCR product was assessed using 1.5% agarose gel electrophoresis (45 minutes at 100 V), and visualised with UV light illumination (Figure 2.5). Only those samples that generated the target product during this initial step were used in subsequent downstream analysis.

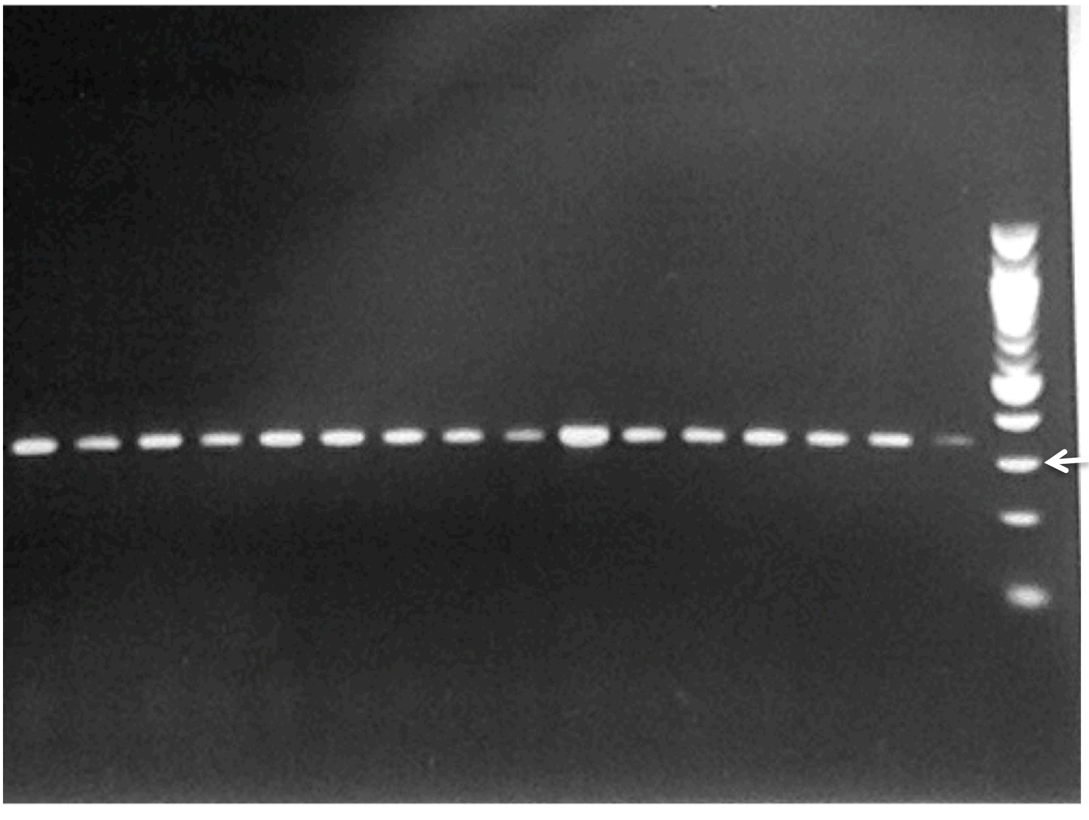


Figure 2.5 Agarose gel of 345 bp PCR product of a reaction using beta-globin primers to assess the integrity of archival DNA. From left: lane 1-16 contains amplified beta-globin product of

archival cSCC DNA (negative control not shown). Lane 17 contains 100 bp DNA ladder, white arrow corresponds to 300 bp, bottom band corresponds to 100 bp. All samples with amplified product were included in further analysis of MCPyV presence.

2.2.6. DNA isolation from blood

DNA was isolated from blood in collaboration with Drs Abha Gulati and Karin Purdlie. Upon collection, 20 ml of blood was mixed with 1.5 ml of EDTA (EDTA concentration 1.8 mg/ml of blood) to prevent blood coagulation. DNA was isolated from 1.5 ml of whole blood using the Nucleon Blood and Cell Culture (BACC) DNA extraction kit (Amersham Biosciences, United Kingdom) following the manufacturer's protocol. Briefly, following deproteinisation using sodium perchlorate, DNA is separated using chloroform and proprietary resin. DNA was eluted in 50 µl of EB Buffer (Qiagen, United Kingdom).

2.2.7. DNA quantification

2.2.7.1. DNA quantification using spectrophotometry

DNA concentration was quantified using Nanodrop® 1000 Spectrophotometer (LabTech International, UK), which measures concentration of DNA (ng/µl) based on the absorbance at UV wavelength of 260nm and the purity of DNA based on the ratio of absorbance at 260nm:280nm. A 260:280 ratio of 1.8 is considered relatively pure for DNA, whereas lower ratios suggest contamination, for example, from protein or phenol. Water was used as a blank reading at the beginning of each set of measurements to clean the pedestal and to ensure the machine was working correctly. Two µl of DNA from each sample were loaded onto the spectrophotometer to determine the concentration and DNA purity. Samples with reading below 10 ng/µl were quantified using Qubit™ Fluorometric Quantitation (Life Technologies, UK).

2.2.7.2. DNA quantification using Qubit™ Fluorometric Quantitation

Sufficient amount of Qubit™ working solution was prepared diluting the Qubit™ dsDNA HS reagent 1:200 in Qubit™ dsDNA HS buffer (Invitrogen, United Kingdom). To prepare standards for calibration, 190 µl of working solution was loaded into two 0.5 ml PCR tubes, and 10 µl of each standard (Invitrogen, United Kingdom) were added to the working solution and vortexed for 5 seconds. 199 µl of the working solution were added to individual 0.5 ml PCR tubes and mixed with 1 µl of sample DNA and vortexed for 5 seconds. The final volume in each tube is 200 ml. Tubes were incubated at room temperature for approximately 2 minutes.

Qubit™2.0. Fluorometer was then calibrated using two standards prepared during the previous step and then all tubes containing sample DNA were read on the fluorometer. The DNA concentration of the sample was calculated by the Fluorometer in ng/µl.

2.2.8. Bisulfite conversion for methylation profiling and bisulfite sequencing

For methylation microarray profiling and bisulfite sequencing, 550 ng or less of genomic DNA was treated with Zymo EZ DNA Methylation-Gold™ Kit (Zymo, United States) following the manufacturer's protocol. Briefly, DNA is denatured at 98°C for ten minutes, followed by 2.5 h incubation at 64°C during which bisulfite conversion takes place. Finally, converted DNA is purified and cleaned up on column, followed by elution in water. This process converts unmethylated cytosines to uracil, leaving methylated cytosines intact. Thereby in sequencing analysis, methylated cytosines are displayed as cytosines and unmethylated cytosines as thymines.

2.2.9. Whole-genome Bisulfite Sequencing

Whole-genome bisulfite sequencing (BS-seq) was carried out in collaboration with Professor Vardhman Rakyan[5]. (BS-seq). This method combining the conversion of unmethylated cytosines to uracil with deep sequencing allows the assessment of methylation at a single nucleotide level. Unlike DNA methylation microarrays, this approach allows the assessment of methylation in repetitive elements that constitute almost a half of all CpGs (Harris et al., 2010). This method has been previously utilised to describe differences in methylation patterns of pluripotency and differentiation genes in embryonic stem cells (Lister et al., 2009).

Pooled DNA from 10 laser-capture microdissected paired normal skin and cSCC samples was used in this experiment. Libraries were prepared with Ovation® Ultralow Methyl-Seq Library Systems (NuGEN, United States) according to the manufacturer's protocol. Overview of the protocol is provided in Figure 2.6.

---

[5] The Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

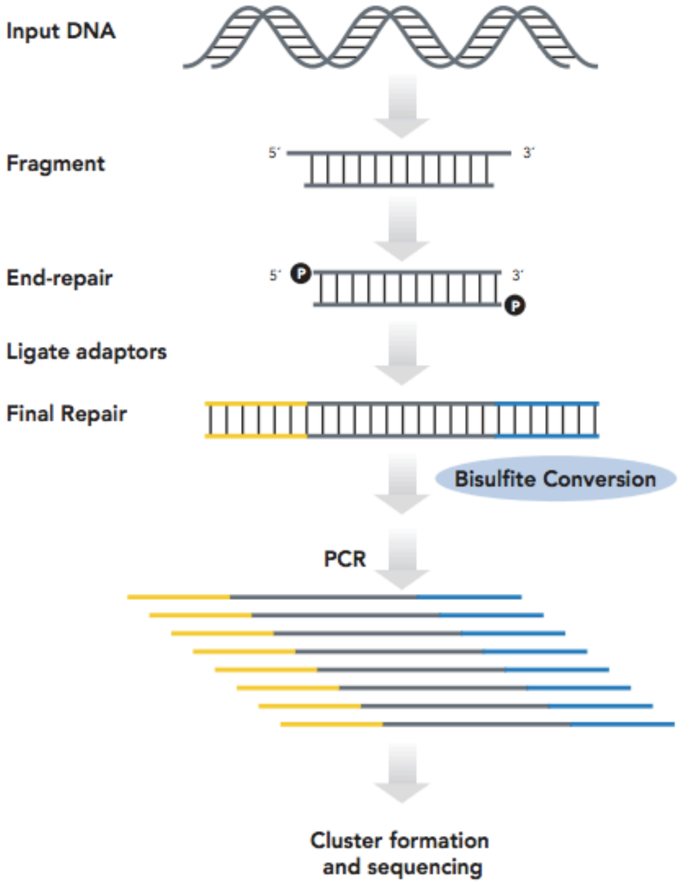


Figure 2.6 Overview of steps involved in the preparation of BS-seq libraries. Figure taken from NuGEN Ovation® Ultralow Methyl-Seq Library Systems protocol.

50 ng of pooled DNA was first sonicated with Bioruptor (Diagenode, United Kingdom) to generate fragments 200-400 bp longs. Fragment size was assessed with Agilent 2100 Bioanalyzer (Agilent Technologies, United States). Agencourt® RNAClean® XP Purification Beads (Beckman Coulter, United States) were then added to the DNA, and DNA bound to RNAClean Up XP beads was removed from the supernatant with a magnet. DNA was then purified with DNA Clean & Concentrator™-5 kit (Zymo, United States). End-repair, ligation of adaptors and final repair was carried out with appropriate buffers (NuGEN, United States), and DNA was bisulfite converted with Zymo EZ DNA Methylation-Gold™ Kit (Zymo, United States). Libraries were then amplified in PCR cycler using Amplification Master Mix with the following cycling conditions:

Initial denaturation	**95°C for 2 minutes**	**1 cycle**
Denaturation	95°C for 15 seconds	12 cycles
Annealing	60°C for 1 minute	
Extension	72°C for 0.5 minutes	
Hold	10°C	∞

Table 2.1 PCR cycling conditions used to amplify BS-seq libraries.

Libraries were then sequenced with Illumina HiSeq 2000 next generation sequencer at Barts and the London Genome Centre. This technology binds randomly fragmented DNA to “flowcells”, an optically transparent surface that encompasses eight separate lanes (one lane can be used for one sample). Each DNA fragment that binds by the adapter to the flowcell is then amplified by bridge-PCR into clusters that contain thousands of copies of the fragment. These clusters are then sequenced by sequencing-by-synthesis technology that adds complementary bases labelled with a fluorescent dye one at a time. Fluorescent dye is then excited with a laser, and image of the fluorescence is acquired before the fluorophore is removed and a new base labelled with a fluorophore is added. Each base is labelled with a unique dye and DNA sequence is then determined by associating respective dyes with paired bases. Once the first-end read is complete, DNA templates are regenerated and sequenced from the other end (each read us sequenced 50 or 100 cycles at each end), which generates paired-end reads. Each lane generates millions of reads, which are then processed *in silico*.

2.2.10. Data analysis of BS-seq data

BS-seq data was analysed in collaboration with Dr Thomas Down using in-house C++ based software BiFast (Lowe et al., 2013). This programme is freely available (https://bitbucket.org/xboxrob/bifast), and combines the previously published BISMARK algorithm (Krueger and Andrews, 2011) with BOWTIE (Langmead et al., 2009). BISMARK algorithm converts cytosine to thymine and guanine to adenosine in the BS-seq reads and aligns those to appropriate segments of the reference genome. The best alignment is then determined based on four different alignment processes run in parallel. BOWTIE, on the other hand, uses a Burrows-Wheeler and full-text minute space index based approach.

Raw BS-seq data was mapped to the reference genome using the BiFast algorithm, and the number of methylated and unmethylated reads at each CpG site was counted.

2.2.10.1. Validation of Illumina450K methylation array data with BS-seq data

For comparison with the Illumina450K methylation data, all CpG sites from the bs-seq dataset lying within 50bp of a differentially-methylated probe detected with Illumina 450K methylation data were then selected if at least 10 reads from both the cSCC and paired normal skin pools aligned properly. Methylation difference between the two pools was then calculated.

2.3. Polymerase chain reaction

2.3.1. Primers list

All primers were ordered from Sigma Aldrich (United Kingdom) and made to a final concentration of 100mM. Merkel Cell Polyoma virus (MCPyV) primers were designed based published work (Feng et al., 2008, Duncavage et al., 2009). The nucleotide sequence of these primers is complementary to various segments of the viral genome (Figure 2.7). MCVPS1 and LT1 primers are complementary to the large T-antigen of MCPyV sequence, LT3 is complementary to the small T-antigen and VP1 hybridises to the VP1 region.

Table 2.2 provides primer sequences, amplicon size and their respective methodological purpose.

Purpose	Gene	Primers	Amplicon size
DNA integrity of DNA isolated from fresh frozen tissue	Beta-globin	KM29(F) GGTTGGCCAATCTACTCCCAGG GH21(R) GGAAAATAGACCAATAGGCAG	345 bp
DNA integrity of DNA isolated from FFPE tissue or archival DNA	Beta-globin	B1 GTGTGCTGGCCCATCACTTT B19 CAAGAAAGCGAGCTTAGTGA	120bp
MCPyV detection	LT1	F:TACAAGCACTCCACCAAAGC R: TCCAATTACAGCTGGCCTCT	439 bp
MCPyV detection	LT3	F:TTGTCTCGCCAGCATTGTAG R:ATATAGGGGCCTCGTCAACC	308 bp
MCPyV detection	VP1	F:TTTGCCAGCTTACAGTGTGG R:TGGATCTAGGCCCTGATTTTT	351 bp
MCPyV detection	MCVPS1	F: TCAGCGTCCCAGGCTTCAGA R: TGGTGGTCTCCTCTCTGCTACTG	109 bp

Table 2.2 Primers used in PCR reactions.

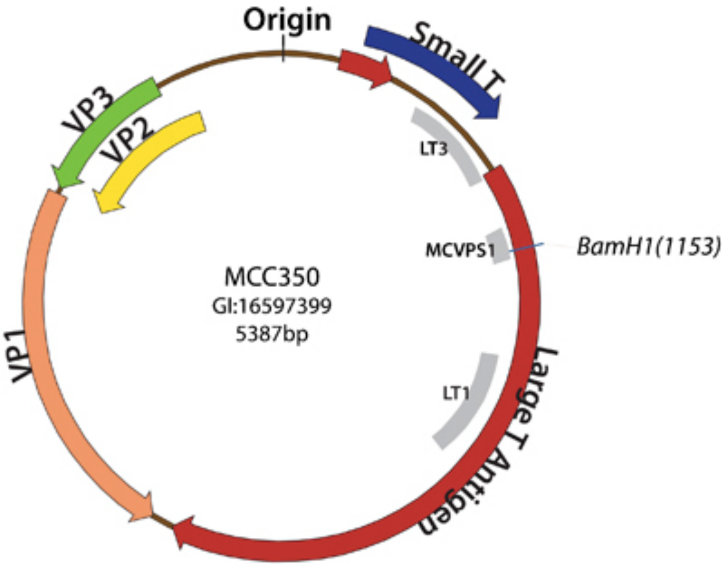


Figure 2.7 Genome of Merkel Cell Polyomavirus. Attachment sites of LT1, LT3 and MCVPS1 primers are shown in grey. VP1 primer binds to the VP1 genomic region. Figure taken from (Duncavage et al., 2009).

Standard PCR reactions were conducted in a 25 or 50 µl total reaction volume using AmpliTaq Gold® DNA Polymerase. Table 2.3 lists components for reaction yielding a total of 50 µl. Cycling parameters for each reaction are listed in the following subsection. PCR amplification was carried out using the MJ Research DNA Engine Dyad thermal cycler (Genetic Research Instrumentation, United Kingdom).

Component	Volume per reaction (µl)	Concentration in Master Mix
Water	27.75	-
10x Buffer mix	5	1x
25 mM MgCl2	6	3 mM
10 mM dATP	1	200 µM
10 mM dCTP	1	200 µM
10 mM dGTP	1	200 µM
10 mM dTTP	1	200 µM
Primer 1	2.5	0.5 µM
Primer 2	2.5	0.5 µM
AmpliTaq Gold DNA Polymerase, LD	0.25`	1.25 Units/reaction
DNA template	2	100 ng/reaction
Total volume	50	-

Table 2.3 Standard PCR sample mix for a total of 50 microliters of final volume.

2.3.1.1. PCR cycling parameters

***Taq* activation**	**95°C for 10 minutes**	**1 cycle**
Denaturation	95°C for 1.1 minutes	60 cycles
Annealing	60°C for 1.1 minutes	
Extension	72°C for 1.1 minutes	
Final extension	72°C for 7 minutes	1 cycle
Hold	4°C	∞

Table 2.4 B1/B19 primers cycling conditions.

***Taq* activation**	**95°C for 10 minutes**	**1 cycle**
Denaturation	95°C for 1 minute	60 cycles
Annealing	55°C for 1 minute	
Extension	72°C for 1 minute	
Final extension	72°C for 10 minutes	1 cycle
Hold	4°C	∞

Table 2.5 KM28/G21 beta-globin primers cycling conditions.

***Taq* activation**	**95°C for 10 minutes**	**1 cycle**
Denaturation	94°C for 0.5 minute	60 cycles
Annealing	50°C for 0.5 minute	
Extension	68°C for 1 minute	
Final extension	72°C for 10 minutes	1 cycle
Hold	4°C	∞

Table 2.6 LT1, LT3 and VP1 primers cycling conditions.

***Taq* activation**	**95°C for 10 minutes**	**1 cycle**
Denaturation	94°C for 0.5 minute	60 cycles
Annealing	55°C for 0.5 minute	
Extension	68°C for 1 minute	
Final extension	72°C for 10 minutes	1 cycle
Hold	4°C	∞

Table 2.7 MCVPS1 primers cycling conditions.

2.3.2. Positive and negative control for PCR

Placental DNA (Sigma Aldrich, United Kingdom) was included in PCR reactions assessing DNA integrity with beta-globin primers to serve as positive control. A total 100 ng of placental DNA were included in each batch of samples amplified with PCR using beta-globin primers to confirm proper run of the reaction.

Plasmid vectors were used as positive control in reactions detecting the presence of MCPyV. Two plasmids were obtained from the University of Pittsburgh Cancer Institute, United States, and the plasmid sequences were homologous to regions that overlapped with the primer sequences. 100 ng of plasmid were used as positive control for each batch of PCR reactions.

Negative control consisted of PCR mastermix with no DNA to exclude contamination with DNA or plasmid and false positive reactions.

PCR products were visualised on agarose gel as described above, and stored at -20°C.

2.4. RNA based techniques

RNA isolation was carried out in a dedicated laboratory area. Prior to RNA isolation, the area and applicable instruments were cleaned with RNase ZAP (Ambion, United States) in order to remove rnases from all surfaces. A centrifuge dedicated to RNA isolation only was used for all steps involving centrifugation. If appropriate, RNA lysis buffer was premixed with β-mercaptoethanol to inhibit rnases present in the sample. On-column DNase I treatment (Qiagen, United Kingdom) was carried out according to manufacturer's instructions in order to minimise contamination with genomic DNA.

2.4.1. Total RNA and miRNA isolation from tissue samples

MiRNAs were isolated using miRNeasy Mini Kit (Qiagen, United Kingdom). Fresh frozen tissue biopsies were cut in half on dry ice. One half was transferred back into the original labelled cryovial, another half was ground using a pre-chilled pestle and mortar and scraped into a pre-chilled 1.5 ml Eppendorf tube. 700 µl of QIAzol lysis reagent (Qiagen, United Kingdom) were added to the tube, vortexed for 1 minutes and transferred into QIAshredder spin column (Qiagen, United Kingdom) placed in a 2 ml collection tube for homogenisation. The samples were then spun for 2 minutes at maximum speed in a dedicated centrifuge. The homogenate was incubated at room temperature for 5 minutes, followed by the addition of 140 μl of chloroform. After 15 seconds of shaking, the sample was further incubated at room temperature for 3 minutes, then centrifuged at 12000g for 15 minutes at 4°C. The aqueous phase was then transferred into a new 1.5 ml Eppendorf tube, mixed with equal amount of pure ethanol and loaded into miRNeasy Mini spin column followed by the standard manufacturer's protocol. Total RNA and miRNA were eluted in 50 μl of water, kept on ice during RNA quantification and stored in -80°C to prevent RNA degradation.

2.4.2. mRNA isolation from cultured cells

2.4.2.1. mRNA isolation from primary cells and established stable cell lines

Cells were pelleted as described in Chapter 2.2.4.1. RNA was extracted from the pellets using RNeasy Mini kit (Qiagen, United Kingdom) following the standard protocol and eluted in 100 μl of sterile water. Briefly, 350 μl of RTL lysis buffer were mixed with the pellet on ice by pipetting and transferred into QIAshredder spin column (Qiagen, United Kingdom) placed in a 2 ml collection tube for homogenisation. The samples were then spun for 2 minutes at maximum speed in a dedicated centrifuge. This homogenate was mixed with equal amount of prechilled 100% ethanol, and mRNA was extracted and cleaned up following on-column treatment with DNAse (Qiagen, United Kingdom).

2.4.2.2. mRNA isolation from fluorescence-activated cell sorting (FACS) isolated cells

Cells were pelleted as described in Chapter 2.2.4.2. RNA was extracted in parallel with DNA isolation using the AllPrep DNA/RNA Micro Kit (Qiagen, UK) following a standard protocol. Briefly, buffer RTL Plus was mixed with beta-mercaptoethanol to inhibit RNAses, and a total of 350 μl were added to the pellet to disrupt cells. This lysate was homogenised using QIAshredder spin column (Qiagen, United Kingdom), placed in AllPrep DNA spin column and spun for 30 seconds at 10 000g. DNA was isolated from this column, and RNA was isolated from the flow-through by mixing it with equal volume of 70% ethanol and followed by standard steps of on-column clean-up, DNase treatment and elution.

RNA was eluted from the column into 50 μl of water.

2.4.3. Laser capture microdissection (LCM) of frozen tissue with subsequent mRNA isolation

This procedure was carried out in collaboration with Dr Sally Lambert. Fresh frozen samples were cut on microtome into 8 μm sections that were mounted on RNase ZAP (Ambion, United States) and UV-light pre-treated PEN slides (described in Chapter 2.2.1). Samples were immediately covered with pre-chilled 70% ethanol for two minutes, stained with pre-chilled cresyl violet and dehydrated in a series of ethanol washes of increasing ethanol concentration.

Following laser-capture microdissection of target cells, isolated cells were collected in Eppendorf tubes containing 350 µl of RTL buffer mixed with beta-mercaptoethanol, and RNA was extracted using RNeasy Micro Kit (Qiagen, United Kingdom) with on-column DNase treatment.

RNA was eluted from the column into 14 µl of water.

2.4.4. RNA quality control

To ensure that RNA is not degraded, 200 ng of total RNA were run on 1.5% agarose gel containing ethidium bromine for 30 minutes at 70 V, and were visualised with UV light illumination.

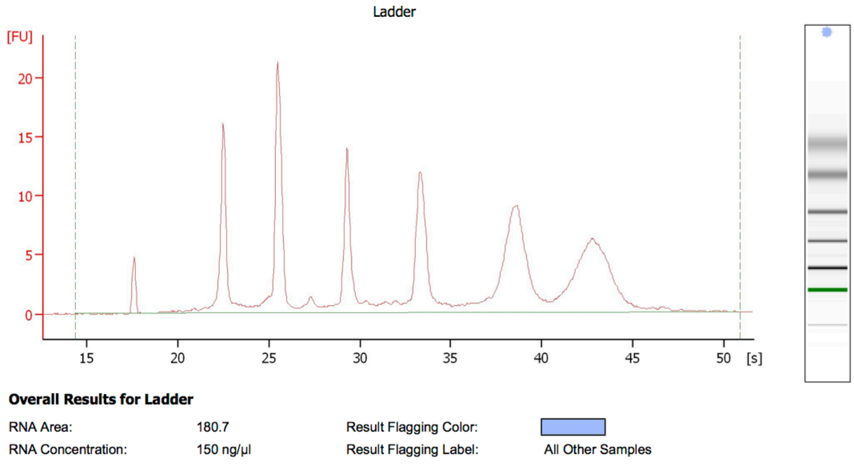
2.4.5. RNA quantification

2.4.5.1. RNA quantification using spectrophotometry

RNA concentration was quantified using Nanodrop® 1000 Spectrophotometer (LabTech International, UK), which measures concentration of RNA (ng/µl) based on the absorbance at UV wavelength of 260nm and the purity of RNA based on the ratio of absorbance at 260nm:280nm. A 260:280 ratio of 2.0 is considered relatively pure for RNA, whereas lower ratios suggest contamination, for example, from protein or guanidine which is rather common occurrence when using column-based kits. Water was used as a blank reading at the beginning of each set of measurements to clean the pedestal and to ensure the machine was working correctly. 1.5 µl of RNA from each sample were loaded onto the spectrophotometer to determine total RNA concentration and RNA purity.

2.4.5.2. RNA quantification and quality control using Agilent Bioanalyser

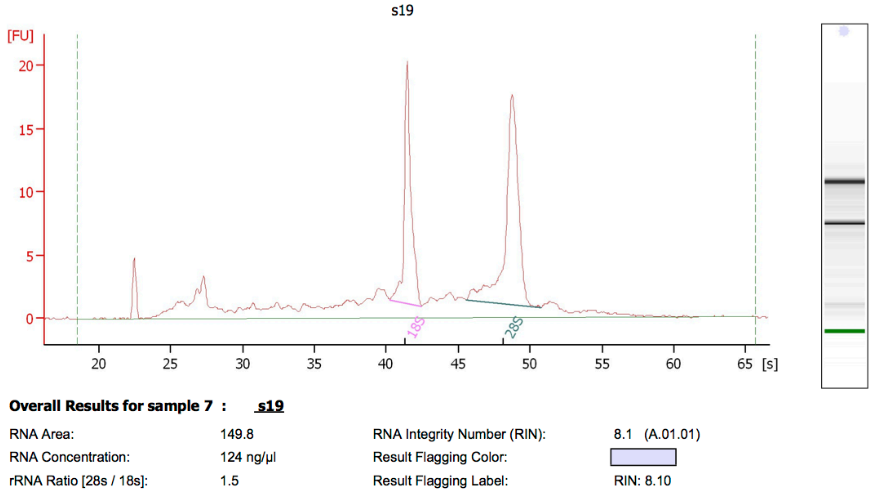
Agilent 2100 Bioanalyser Chipreader (Agilent, United Kingdom), uses electrophoresis to quantify DNA, RNA or protein in multiple samples simultaneously using a chip. Each chip consists of 16 wells, 12 of which are used for samples, 1 for ladder and 3 for matrix reservoir. Samples move from the well through a series of microchannels filled with fluorescent dye and polymer due to electric current, followed by electrokinetic injection into separation channel that separates the sample based on their molecular weight, and provides a scanned image of the components (Figure 2.8). In addition, the software provides a proprietary RIN value that describes RNA quality. For transcription profiling, RIN value of 8 or more is desirable.



Overall Results for Ladder

RNA Area: 180.7
RNA Concentration: 150 ng/µl
Result Flagging Color: 🟦
Result Flagging Label: All Other Samples

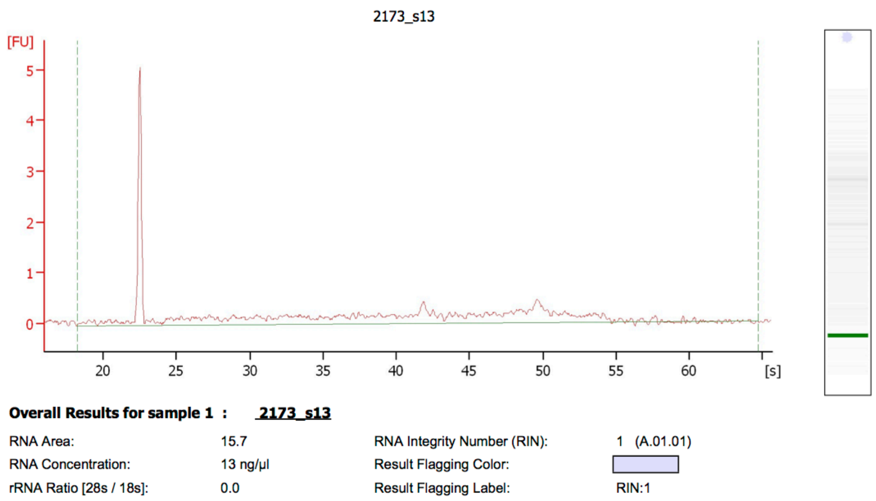
Figure 2.8 Agilent Bionanalyser RNA 6000 Nano ladder. This ladder contains a mixture of six RNA fragments of known size and concentration (150 ng/µl), and is used as a reference for the data analysis of experimental samples. RNA peaks of experimental samples are compared to the ladder fragments to determine RNA concentration.



Overall Results for sample 7 : s19

RNA Area: 149.8
RNA Concentration: 124 ng/µl
rRNA Ratio [28s / 18s]: 1.5
RNA Integrity Number (RIN): 8.1 (A.01.01)
Result Flagging Color: 🟪
Result Flagging Label: RIN: 8.10

Figure 2.9 Agilent Bioanalyser output for a sample of RIN 8.1. This figure shows two distinct ribosomal peaks corresponding to 18S and 28S and a mostly flat baseline between the internal marker (left-most peak) and 18S. A smaller peak next to the internal marker corresponds to 5S. These parameters represents a high-quality RNA isolated from a tissue sample.



Overall Results for sample 1 : 2173_s13

RNA Area:	15.7	RNA Integrity Number (RIN):	1 (A.01.01)
RNA Concentration:	13 ng/µl	Result Flagging Color:	▨
rRNA Ratio [28s / 18s]:	0.0	Result Flagging Label:	RIN:1

Figure 2.10 Agilent Bioanalyser output for a sample of RIN 1. Peaks corresponding to 18S, 28S or 5S are almost completely absent. The most prominent peak corresponds to internal marker. Such RNA is likely highly degraded.

2.4.6. cDNA synthesis for RT-PCR

Three hundred nanograms of total mRNA were used for first-strand cDNA synthesis with qScript™ cDNA SuperMix (Quanta BioSciences, United States). RNA template was mixed with qScript cDNA SuperMix (5X) and RNase-free water in 0.2 µl tubes on ice to make a total of 20 µl per reaction. Each tube was briefly vortexed and centrifuged to collect the mixture on the bottom of the tube. Samples were then incubated in PCR cycler as follows:

Time	Temperature
5 minutes	25°C
30 minutes	42°C
5 minutes	85°C
Hold	4°C

Table 2.8 Cycling conditions for first-strand cDNA synthesis.

2.4.7. Primer design

Primer for reverse-transcriptase – polymerase chain reaction (Q-PCR) were designed with Primer3 software (Untergasser et al., 2012). Ensembl genome browser was then used to retrieve exons sequence, and the sequences of two adjacent exons were entered into Primer3. Primers that yield product between 85-150 bp with melting temperature (Tm) close to 60°C were selected for the experiment. Complementarity of selected primer to the proper gene were confirmed with USCS genome browser. All primers were obtained as custom oligonucleotides from Sigma Aldrich (United Kingdom), and resuspended in distilled water to make 100 nM solution. Primers are listed in Table 2.9.

Primer	Forward sequence (5'-3)	Reverse sequence (5'-3)	Annealing

			temperature
Beta 1-integrin	CAAAGGAACAGCAGAGAAGC	ATTGAGTAAGACAGGTCCATAAGG	59°C
CD44	GAAGAAAGCCAGTGCGTCTC	CACGTGGAATACACCTGCAA	59°C
ESA	CCTCCACGTTCCTCTATCCA	AAGCAGTTTACGGCCAGCTT	59°C
GAPDH	AATGAAGGGGTCATTGATGG	AAGGTGAAGGTCGGAGTCAA	59°C

Table 2.9 Primers used for QPCR reactions.

2.4.8. Reverse-transcriptase – polymerase chain reaction (Q-PCR)

QPCR was carried out with MESA Blue qPCR MaterMix Plus or SYBR© Assay (Eurogentec, United Kingdom). All reagents were thawed on ice and mixed by repeated inversion prior to the experiment. 6.5 μL of 2x reaction buffer were mixed with 2.5 μL of 100 nM reverse and forward primers, two microliters of the final cDNA product generated with the previous reaction to make a total of 13.5 μL reaction volume. Master mix containing all components except for the cDNA template was carefully pipetted into wells of 96-well PCR plate according to a pre-designed layout, cDNA was added at the end and the plate was sealed. All procedures were carried out in laminar flow hood with no illumination to prevent contamination and unnecessary exposure to light. All reactions were performed in triplicate, and a triplicate blank control containing no cDNA template was included on each 96 well PCR plate.

QPCR data were obtained with the AB7500 Fast Realtime PCR System (Applied Biosystems, United Kingdom) and extracted with the system's software. Dissociation curve was obtained for each experiment. Cycling condition for the three-step protocol are describe in Table 2.10.

Meteor*Taq* activation	**95°C for 5 minutes**	**1 cycle**
Denaturation	95°C for 15 seconds	40 cycles
Annealing	60°C for 20 seconds	
Extension	72°C for 40 seconds	

Table 2.10 Cycling conditions for three-step qPCR protocol.

2.5. Antibody-based techniques

The following mouse monoclonal antibodies were used in subsequent antibody-based techniques:

Antibody	**Source**	**Clone**	**Conjugate**
Beta-1 integrin	Abcam	B-D15	Fluorescein isothiocyanate (FITC)
CD44	BD Pharmingen™	G44-26	Phycoerythrin (PE)
ESA	Miltenyi Biotec	HEA-125	Allophycocyanin (APC)

Table 2.11 Antibodies used in FACS and immunocytochemistry

2.5.1. Fluorescence-activated cell sorting

Cells were cultured in a monolayer until reaching 70% confluency. After removing the media and washing the cells in PBS, cells were detached with Accutase® (Sigma Aldrich, United Kingdom) to minimise damage to surface epitopes. Cells were incubated with 5-10 mL Accutase at 37°C in 5% CO2 incubator for 3 to 5 minutes. The maximum incubation time was 10 minutes. Accutase was then deactivated with warm media, and 25 ml of cold PBS was added to the cellular suspension. Cells were transferred into 50 mL Falcon tube and centrifuged at 1500 rpm for 5 minutes at 4°C. The supernatant was then removed and cells were resuspended in 0.5 ml of cold PBS. 5 µl of each antibody were added to the cells (1:100 dilution) and were incubated at 4°C in darkness for 15 minutes. Cells were then diluted with cold PBS and centrifuged at 1500 rpm for 5 minutes at 4°C. Supernatant was removed and the cell pellet resuspended in 300-500 µl of cold PBS. Cells were passed through 100 um Cell strainer (BD Sciences, United Kingdom) to 50 mL Falcon tube. 10 uL of DAPI nuclear dye (Sigma Aldrich, United Kingdom) was added to the suspension to exclude dead cells. Cells were then transferred into polypropylene tubes and 3% triple-positive and negative cells were sorted with the BD FACSAria™ I machine (BD Biociences, United Kingdom) into pre-chilled culture media. Immediately after isolation of cells, these were transferred into 50 ml Falcon tube containing cold media and centrifuged at 1500 rpm for 5 minutes at 4°C.

Cell pellet was then resuspended in 2 ml fresh, cold media and transferred into 6-well plate containing 2 ml of chilled pre-cultured media. Combining fresh and pre-cultured media in 1:1 ratio has significantly increased the viability of sorted cells. Cells were placed to 37°C in 5% CO2 incubator and the media was allowed to gradually warm up in the incubator to minimise stress to cells.

2.5.1.1. Compensation for PE and FITC

Because the emission spectrum of PE and FITC partially overlaps (Figure 2.11), compensation was carried out with Becton Dickinson Compensation Beads (BD Biociences, United Kingdom). These beads show either strong positive fluorescence peak from the antibody, or a completely negative peak. Unlabelled cells can create false positive signal by autofluorescence, while the beads completely block the fluorescent signal. Compensation is then computed by subtracting negative and positive populations using the compensation function in BD FACSDiva Software.

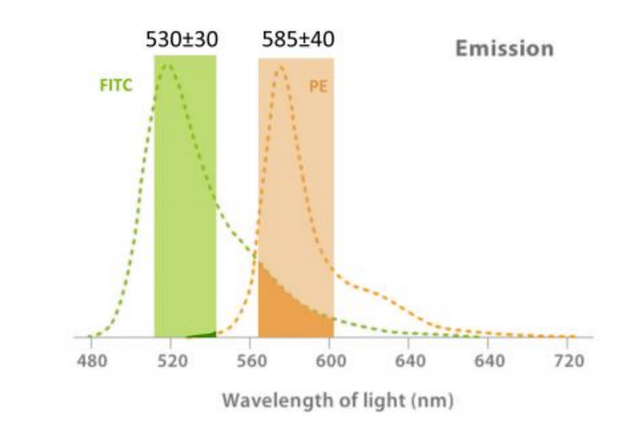


Figure 2.11 Overlap of FITC and PE emission spectrum. PE is an orange fluorophore, yet the emission spectrum of the green fluorophore FITC does slightly overlap into the emission spectrum of PE. In effect, samples labelled with FITC will appear positive for PE to some extent, and thereby compensation is required if these two fluorophores are combined in an experiment. Figure taken from http://www.lifesci.dundee.ac.uk/cast/flow-cytometry-core-facility/techniques-4.

2.5.2. Fluorescent immunocytochemistry

Stable cSCC cell lines were cultured to 50-70% confluency. Cells were detached with Accutase©, mixed with warm media and centrifuged in 50 mL Falcon tube at 1500 rpm for 5 minutes at room temperature. Cells were then resuspended in 5 mL of warm media, and 200 uL were mixed with 300 uL of 0.4% Trypan blue solution, and 20 uL were transferred to a haemocytometer chamber and cells were counted in four corner squares and the centre square under a light microscope. Cells were then seeded onto sterile Vectashield coverslips (Vector Laboratories, United States) in a 12-well plate at $5x10^4$ density, and allowed to attach overnight in an incubator. Media was replaced with PBS, and cells were washed twice for 5 minutes. Cells were then fixed with 1 mL of 4% PFA for 10 minutes at room temperature, and subsequently washed twice with PBS for 5 minutes each.

Cells were then permeabilised with 0.1% Triton X-100 in PBS (PBS-T) for 5 minutes at room temperature, and washed twice with PBS for 5 minutes each. Cells were then blocked with 500 µl of 5% goat serum dissolved in PBS for 1 hour at room temperature, and incubated with fluorophore-labelled primary antibody at 4°C in darkness overnight. Primary antibody was diluted in 5% goat serum in PBS 1:250.

Primary antibody was then removed and cells were washed three times with PBS (5 minutes each wash). DAPI stain (500 ug/mL) was added to each well, and after 10 minute-incubation, cells were washed again twice. Coverslips were then carefully turned over and mounted onto a glass slide with Vectashield Mounting Medium (Vector Laboratories, United States). Slides were immediately visualised with 510 Meta inverted confocal microscope (Zeiss, Germany), and stored at 4°C in darkness.

2.6. Microarray profiling

2.6.1. Biological material, microarrays and data analysis

Copy-number variation in clinical tissues (normal skin, AK) was detected with Genome-Wide Human SNP Array 6.0 (Affymetrix, United States). Transcriptional profiling of normal skin, AK and cSCC clinical samples was carried out with Human Genome U133 Plus 2.0 Array, profiling of cSCC cell lines, cultured primary human keratinocytes and cancer stem cells with HumanHT-12 v3 Expression BeadChip. Methylation profiling of stable cSCC cell lines and primary human keratinocytes was carried out with Infinium HumanMethylation27 BeadChip, profiling of skin, AK and cSCC clinical specimen, cSCC cancer stem cells and MCPyV positive and negative archival cSCC DNA was carried out with Infinium HumanMethylation450 BeadChip Kit. MiRNA profiling of cSCC samples and paired normal skin was conducted with miRCURY LNA™ microRNA Array (7'th GEN).

Details regarding biological samples hybridised to each array, array descriptions and specifications and detailed analytical steps are provided in respective chapters.

2.6.2. Analytical approaches common to the processing of all datasets

Microarray data analysis was conducted in R and Bioconductor (Gentleman et al., 2004). All microarrays were initially subjected to stringent quality control (details are provided in respective chapters), and only those chips that passed the predefined quality control criteria were subsequently normalised. Unless otherwise stated, all data were normalised using quantile normalisation. The principle of this algorithm is to unify the distribution of probe levels across all samples in the dataset (Bolstad et al., 2003). Clustering of microarray data was achieved using Euclidian distance in the *gplots* package (http://CRAN.R-project.org/package=gplots). Specifics and details of the analytical approaches applied are included below and in respective chapters where appropriate.

2.6.2.1. Clustering analysis

Hierarchical clustering of genome-wide data was performed to detect profile correlation between samples using vsn (http://www.bioconductor.org/packages/2.3/bioc/html/vsn.html) package on the normalised data matrix. Hierarchical clustering algorithm with Pearson correlation metric was applied for this purpose. Furthermore, clustering was quantified and further validated using pvclust package (http://www.is.titech.ac.jp/~shimo/prog/pvclust/) in R. This package calculates the statistical significance of observed clusters. P-values are calculated for each via multiscale bootstrap resampling, and the p-value ranges from 0 to 1, indicating the strength of support for the cluster. The level of 0.95 was set as α level of significance for the existence of a given cluster.

2.6.2.2. Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) searches for common biological function or pathway regulation of given gene or probeset. Using in-house Pearl scripts, each probe was annotated with respective pathways it interacts with according to the KEGG database and a chi-square test was performed on differentially methylated probes using all probes as the background. An analogous approach was applied to Gene Ontology enrichment analysis: Probes were annotated with respective GO terms by interrogating GO database (www.geneontology.org) and chi-square test was applied in an identical manner. Molecular Signatures Database (MsigDB) was used to test for cytogenetic location enrichment (Liberzon et al., 2011). α-level for enrichment analysis was set at pvalue<0.05.

2.7. Tissue culture techniques

2.7.1. Tissue culture of primary human keratinocytes and cSCC

Primary human keratinocytes (PHK) were derived from clinical samples or obtained as a gift from Drs A. Bahta, P. A. Costa and R. Hannen[6].

PHK were cultured in DMEM/Ham's F12 medium prepared in a 3:1 ratio by a commercial tissue culture company (Gibco, United Kingdom). The medium was supplemented with 10% fetal calf serum (FCS, Biosera, United States), L-glutamine (Sigma Aldrich, United Kingdom) and RM+ (for composition, see Table 2.12).

Reagent	Final concentration in culture medium
Hydrocortisone	0.4 μg/ml
Cholera toxin	10^{-10} M
Transferrin	5 μg/ml
Liothyronine	2×10^{-11} M
Adenine	1.8×10^{-4} M
Insulin	5 μg/ml
Epidermal growth factor (EGF)	10 ng/ml

Table 2.12 Composition of RM+ reagent.

Upon receiving a clinical tissue sample (normal skin biopsy or cSCC biopsy tissue), the specimen was placed in trypsin at 37°C for 1 hour, and primary keratinocytes or malignant keratinocytes

---

[6] All members of Department of Cutaneous research, The Blizard Institute, Queen Mary, University of London, UK

were scraped off with a sterile scalpel into 100 ml petri dish in a laminar flow hood. Approximately 500,000 3T3 mouse fibroblasts grown in T25 culture flasks in DMEM medium supplemented with 10% FCS were treated with mitomycin C for 3 hours to inhibit their proliferative potential. Scraped cells were added to those flasks, and grown in DMEM/Ham's F12 medium until confluent (3-7 days). Prior to PHK/primary cSCC cells collection, T25 flasks were cleared of 3T3 fibroblasts by treatment with versine reagent (0.02% EDTA in PBS, Gibco, United Kingdom).

2.7.2. Tissue culture of primary cSCC cell lines

Primary cutaneous cSCC cell lines derived from patient tumours collected at Barts and The London Dermatology clinic were obtained in cryovials from long-term storage in liquid nitrogen tanks. Vials were quickly thawed (< 1 minute) in a 37°C water bath, added to 50 ml Falcon tube containing 20 ml pre-warmed DMEM/F12 medium supplemented with 10% FCS and centrifuged at 1500 rpm for 5 minutes. The supernatant was subsequently removed, and the pellet was resuspended in 7 ml of 10% FCS-supplemented DMEM/F12 medium in T25 culture flask and placed in 37°C, 5% CO2 incubator. Primary cSCC cell lines were grown without 3T3 feeders.

2.7.3. Collection and long-term storage of cultured cells

Prior to cell collection, culture medium was removed, cells were carefully washed with PBS and incubated with 0.05% trypsin and 0.02% EDTA solution in PBS in 37°C, 5% CO2 incubator until cells detached from the flask. Trypsin was inactivated by the addition of culture media, and cell suspension was then transferred into 15 ml Falcon tube and following 5 minutes of centrifugation at 1500 rpm and supernatant removal, either resuspended in culture media and re-seeded for further tissue culture, immediately frozen for isolation of nucleic acids, or resuspended in 1.2-1.5 ml of freezing media (70% DMEM/F12 tissue culture medium, 20% FCS, 10% DMSO), and this suspension was transferred into a cryovial. Cryovials were placed in a Nalgene Mr. Frosty Freezing Container that provides -1°C/min cooling (Fisher Scientific, United Kingdom) when placed in -80°C freezer. Cryovials were then transferred to liquid nitrogen tanks for long-term storage.

2.7.4. Culture of embryonic stem cells

Human embryonic stem cells (hESC) were cultured in collaboration with Dr Cleo Bishop[7]. HESC H9 cells were cultured under licence from Stem Cell Technologies. HESC cells were cultured on a monolayer of mitomycin-treated mouse embryonic fibroblasts (MEF) in hESC media. The hESC medium was prepared fresh every two weeks, and changed every day. The composition of hESC media is listed in Table 2.13. PBS, beta-mercaptoethanol and L-glutamine were mixed separately

---

[7] Department of Cutaneous Research, The Blizard Institute, Queen Mary, University of London, United Kingdom.

and prepared fresh each time. The medium was then filtered through 0.22uM filter and stored at 4°C for a maximum of 14 days.

Solution	**Volume for 250 ml**
DMEM/F12 Medium	200 ml
Knockout Serum Replacer (KO)	50 ml
Non-Essential Amino Acids (100 X)	2.5 ml
L-Glutamine (100 X)	2.5 ml
PBS with Ca^{2+} and Mg^{2+}	2.5 ml
Beta-Mercaptoethanol	3.5µl
bFGF (2 ug/ml)	500µl

Table 2.13 Composition of hESC medium (total volume of 250 ml).

Prior to hESC passaging, mitomycin C-inactivated MEFs were plated onto gelatin-coated wells of 6-well plate (1 mL of 0.1% sterile aqueous gelatin solution per well) and allowed to attach in MEF media overnight. Composition of the MEF media is provided in Table 2.14.

Reagent	**Volume**
DMEM high glucose (without L-glut)	500mL
Fetal Bovine Serum	50mL
100 x L-Glutamine	5mL
Sodium Pyruvate	5mL
Non-essential amino acids	5mL
β-mercaptoethanol	3.5uL

Table 2.14 Composition of the MEF medium.

MEFs were kept as a feeder layer for a maximum of 10 days, and were inspected visually the next day after seeding to confirm proper morphology before hESC were passaged.

Prior to passaging, hESC colonies were inspected under the microscope and differentiated colonies were manually removed from culture. Then 10-20 uL of fresh Dispase were added to each well, and cells were incubated at 37°C for a maximum of 30 minutes. Plates were examined under the microscope every 5-10 minutes, and once colonies dislodged, these were transferred into 10 mL Falcon tube and allowed to settle by gravity. Cells were then washed with 5 ml of warm medium three times. MEF medium was replaced with warm hESC medium in new MEF-covered plates, and washed hESC colonies were resuspend in 5mL hESC medium and gently disrupted with a 5mL pipette. Cells were then seeded into new wells, gently swirled to ensure even distribution of cells and allowed to settle overnight. Medium was changed the next day.

2.7.5. Cryopreservation of hESC

Only colonies of cells of proper morphology maintained in fully established culture were cryopreserved. Colonies were cryopreserved 7-8 days post-passaging to ensure the colonies reached their maximum size and the best recovery upon thawing. 10-20uL of fresh Dispase were added to the media and incubated at 37^{o}C until the colonies fully detached. Colonies were then transferred to a 50 ml or 15 ml Falcon tube, depending on the number of colonies. Cells were then allowed to settle by gravity or by centrifugation at 300rpm for 1 minute at room temperature.

Cells were then washed twice with 5mL hESC medium as described in the general hESC passaging step. HESC colony pellet was then resuspended in 5mL of hESC medium and centrifuged at 300rpm for 5 minutes at room temperature. The supernatant was carefully removed and the pellet was then resuspended with 500uL per well of hESC medium.

10 mL of CRYO medium was prepared by mixing 6 mL of Knockout serum, 2 mL of normal hESC medium and 2 mL of DMSO. 500 uL of this mixture per well were added drop by drop to to resuspended hESC colonies. 1 mL of the 1:1 medium/CRYO medium mixture was added to cryovials. Cryovials were placed in a Nalgene Mr. Frosty Freezing Container and placed in -80°C freezer. Cryovials were then transferred to liquid nitrogen tanks for long-term storage.

2.7.6. Thawing of cryopreserved hESC

The average recovery of hESC after cryopreservation is 1-10%. Wells were coated with 0.1% gelatin and MEFs plated one day before thawing of hESC. On the day of thawing, MEF medium was replaced with warm hESC medium and MEFs were incubated at 37^{o}C. Cryovials were thawed in water bath, the the content was transferred into 4 ml of warm media in a 15mL Falcon tube. Thawed colonies were allowed to settle by gravity in water bath to keep colonies warm. The supernatant was carefully removed and the pellet was washed with 5 mL hESC medium. Medium was removed from wells containing MEFs, and colonies were added drop by drop to each well. The plate was gently swirled to ensure even distribution of cells across all wells, and placed at 37^{o}C into the incubator. Colonies were allowed to settle overnight, and medium was changed the next day.

2.8. Standard buffers and reagents

DMEM/Ham’s F12 medium composition

3 volumes of Ham’s F12 medium, 1 volume of Dulbecco's modified eagle's medium (DMEM), fetal calf serum (10%), 4mM L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin.

10x TBE buffer (1 L)

108 g Tris-base, 55 g boric acid, 200 ml 0.5M EDTA (pH 8.0), distilled water to 1000 mL.

1x TBE buffer (1 L)

100 mL of 10x TBE buffer were mixed with 900 mL of distilled water.

10x PBS

80 g of NaCl, 2 g of KCl, 14.4 g of Na2HPO4, 2.4 g of KH2PO4 are mixed with 800 mL distilled water.

0.1% Triton X-100 solution in PBS (PBS-T)

10 mL of 10x PBS were added to 90 mL of distilled water and 0.1 ml of Triton X-100 were added to the mixture.

4% Paraformaldehyde

200 mL of distilled water heated to 60°C, 10 g of laboratory-grade paraformaldehyde, 50 μl of 10N sodium hydroxide, 25 mL of 10X PBS, distilled water to 250 mL.

Orange G Loading Dye

100 mg Orange G, 7.5 mL glycerol, distilled water to 50mL.

10x Trypsin-EDTA

0.5 g of powdered Trypsin, 80 mL of distilled water, 10 mL of 10x PBS and 1 mL of 0.5 M EDTA solution, distilled water to 100 ml.

DAPI stock solution (1 mg/ml)

1 mg of 4′,6-dimidino-2-phenylindole (DAPI) is dissolved in 1 ml of 1x PBS. Stock solution is wrapped in aluminium foil and stored at -80°C for long-term storage, and at 4°C in darkness for short-term storage.

MTT 10x stock solution

25 g of 3-[4,5-Dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide (MTT) were dissolved in 50 mL of 1x PBS to make 5 mg/mL 10x stock solution. 5 mL aliquots of the stock solution were made in 15 ml Falcon tubes, wrapped in aluminium foil and frozen at -20°C for long-term storage.

Trypan blue 0.4% solution

4 mg of Trypan blue powder were dissolved in 80 mL of 1x PBS and warmed up with stirring until fully dissolved. 1x PBS is added to 100 mL.

3. Somatic Copy Number Alterations and Gene Transcription Profiling of Normal Skin and Actinic Keratoses

3.1. Introduction

A plethora of evidence has established that malignant progression is accompanied by genomic instability and subsequent genomic changes. Such genomic changes lead to transcriptional differences between normal, premalignant and malignant tissues, promoting malignant progression (Bayani et al., 2002). As described previously, up to 60% of cSCC are reported to arise from precursor actinic keratoses (AK) – a common clinical manifestation of intraepithelial keratinocyte dysplasia, yet the genetic basis of AK is practically unknown.

Microarray expression profiling is a well-established technique capable of reporting a genome-wide transcriptome, and thus provide novel insights into cellular processes at an mRNA level (Conway and Schoolnik, 2003). High-density oligonucleotide microarrays were originally designed to capture single-nucleotide polymorphisms, but can be applied in detection of both genotype and copy number variation data (Bignell et al., 2004). SNP is a variation of a single nucleotide in a DNA sequence that may occur in both coding and non-coding regions of DNA, and is present in at least 1% of the general population. The most common cause of SNP is point mutation leading to a nucleotide substitution. SNP may or may not effect gene expression or the coded protein structure or function, depending on its character (synonymous versus missense and nonsense) and location in the genome (gene, regulatory sequence, intergenic region). For example, synonymous SNP do not change the protein structure, since the coded amino acid is identical due to codon degeneracy. Figure 3.1 illustrates SNP in the genome.

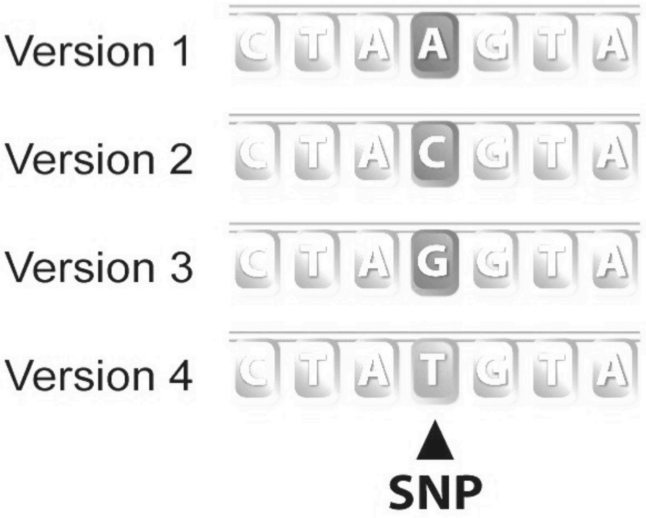


Figure 3.1 Illustration of Single Nucleotide Polymorphism (SNP). Each version represents a single nucleotide substitution in a DNA sequence, and thereby up to four SNP versions may exist in a population, with different percentual distribution for each nucleotide. If SNP occurs outside of a gene sequence, it is categorised as "linked SNP" and has no effect on protein production. SNPs in the regulatory sequences of genes, on the other hand, can alter gene expression levels, while SNPs in the gene coding regions can alter the amino acid sequence of the gene protein. Figure taken from http://learn.genetics.utah.edu/content/pharma/snips/.

Unlike SNP, Copy-number variation (CNV) is a form of structural change that alters the amount of genomic material in a cell by either loss or gain of a region of DNA (copy number imbalance), or leads to copy-neutral loss of heterozygosity (LOH), also known as uniparental disomy in which two copies of a genomic region inherited from one parent exist in the absence of a copy from the other parent. Seemingly phenotypically neutral CNV are common in the human population and have been linked with evolution and phenotypic diversity (Henrichsen et al., 2009), while uniparental disomy has been linked with cancer development in humans due to aberrant imprinting or homozygosity of mutated or methylated genes (Fu et al., 2008). For the purpose of this work, the term "somatic copy number alteration" (SCNA) is used to refer to an alteration in the number of copies of a region of DNA (loss or gain) that is acquired in somatic cells. CNV commonly refers to a different copy number of a germline, rather than somatic DNA (Moubayed et al., 2007).

SNP arrays allow the detection of gains and losses of small genomic regions, LOH and uniparental disomy. Oligonucleotide probes hybridise with fluorescence-labelled DNA fragments and the intensity thereby reflects the number of copies of given DNA regions. Additionally, SNP arrays are able to distinguish parental alleles based on SNPs.

The purpose of this experiment was to find shared genomic changes that underpin AK onset and progression to cSCC and to define the extent to which these genomic changes determine transcriptional changes in their respective lesions. Any discrepancies between these profiles might indicate a potential role for epigenomic changes.

3.2. Materials and Methods

3.2.1. Patients and tissue samples

Skin, AK and tumour samples were collected between 1995-2007 in the dermatology clinics at Barts and the London NHS Trust (London, United Kingdom) by Dr Abha Gulati[8] or Professor Harwood[9] from consenting study participants (clinical details are provided in Appendix 18): fresh frozen biopsies of normal skin, AK and cSCC were obtained at the time of surgical excision. Tumour samples and a subset of AK biopsies were immediately snap-frozen in liquid nitrogen. Two 4 mm biopsies were obtained from skin samples, one of which was also immediately snap-frozen in liquid nitrogen, and the other was split in half: one half was fixed in formalin and embedded in paraffin, the other half was submerged in RNALater RNA preserving agent (Qiagen, United Kingdom) for up to 4 hours prior to freezing in liquid nitrogen. Venous blood was drawn from patients for genotyping. All samples were stored at -80°C until processed. All participating patients provided informed consent in accordance with ethical approval from the East London and City Health Authority local ethics committee.

3.2.1.1. Sun-exposed, non-sunexposed skin and actinic keratoses

Sun-exposed skin was defined as a sample collected from an area usually not covered by commonly worn clothes, such as the outer upper arm or scalp (see Figure 3.2). *Non-sunexposed* skin, on the other hand, was collected from an area usually not exposed to sunlight, such as inner upper arm and buttocks. AKs were collected from confluent areas of keratinocyte intraepithelial neoplasia (KIN). AKs were defined clinically and confirmed histopathologically.

---

[8] Cutaneous Department, Blizard Institute of Cell and Molecular Science, Queen Mary, University of London

[9] Barts and the London School of Medicine and Dentistry; Blizard Institute of Cell and Molecular Science, Queen Mary, University of London

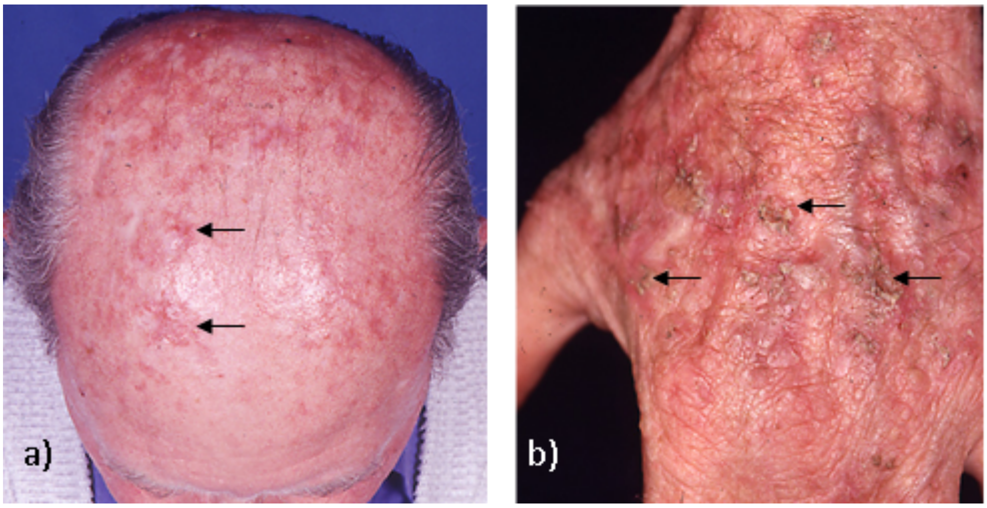


Figure 3.2 Broad areas of keratinocyte intraepithelial neoplasia. Arrows point to individual lesions. Photos prepared in collaboration with Dr Abha Gulati.

Each skin and AK sample was evaluated by a dermatopathologist in a blinded manner to confirm its histological structure.

DNA isolation from blood, skin and AK samples was conducted in collaboration with Dr Abha Gulati. DNA was isolated from those samples either with DNEasy Micro or Mini kit (Qiagen, United Kingdom) following laser-capture microdissection of the sample (described in Chapter 2) or by microdissection under light microscope, during which the epidermis was separated using a scalpel.

If the total DNA yield was below 25 ng/ µl, DNA was amplified using either Illustra Genomiphi v2.0 DNA Amplification kit (GE Healthcare, USA) or Repli-g Mini kit (Qiagen, United Kingdom) following the manufacturers' protocols.

3.2.2. Affymetrix GeneChip Human Mapping 250K Nsp SNP Array

Affymetrix GeneChip Human Mapping 250K Nsp SNP Array (Affymetrix, USA) contains 250 000 oligonucleotide sequence probes complementary to corresponding genomic regions. Prior to hybridisation to the microarray, DNA is digested with restriction enzyme Nsp I and ligated to Nsp-specific adaptors. Adaptor specific primers were then used to amplify the adaptor-linked DNA during a PCR. The PCR products are purified using ultra-free MC filtration column (Millipore, UK) and finally, purified PCR products are fragmented using DNase I, labelled with biotin-labelled reagent and hybridised to the array. The array is then washed, stained with GeneChip Fluidics Station 450 (Affymetrix, USA) and scanned with a GeneChip Scanner 30007G (Affymetrix, USA). Array processing was carried out by Dr Tracy Chaplin[10].

---

[10] Cancer Research UK, Medical Oncology Laboratory, Barts and the London School of Medicine and Dentistry, Queen Mary, University of London.

Data acquisition from the SNP array consisted of image analysis by GeneChip Operating Software (Affymetrix, USA) generating probe cell intensity data. SNP allele calls were generated by GeneChip Genotyping Analysis Software (Affymetrix, USA). The raw values were then preprocessed and normalised in Bioconductor in collaboration with Fios Genomics[11]. I analysed the data using the following packages: ggplot2, gviz, and glm regression analysis function in R. Numeric coding of SCNA is the following: CN0 = homozygous deletion, CN1=heterozygous deletion, CN2=uniparental disomy, CN3=heterozygous gain, CN4=double gain.

Size of SCNA was divided into 5 categories (XS, S, M, L, XL) based on the following cut-off criteria: XS < 10K bp, S < 100 K bp, M < 1 MB bp, L < 10 MB bp, XL > 10 MB.

3.3. Results

3.3.1. Somatic copy number alteration in normal skin and actinic keratoses

3.3.1.1. Experimental design and comparison strategy

SNP array profiling data from a total of 38 patients were included in the study. Each of those patients provided a blood sample, AK, sun-exposed and non-sunexposed skin biopsy. Only arrays that passed quality control and samples with histological verification of tissue characteristics were included in the analysis. I focused on somatic mutations potentially relevant to non-melanoma skin onset, specifically progression from normal skin to premalignancy (AK).

Data variables included: chromosome, SNP coordinates (start, end), size in bp, copy number (CN0-4), raw data filename, SNP name, chip type, scan date, immune status, tissue type, gender, patient ID, extraction kit, histology, and microdissection type.

In order to exclude genomic changes due to germline mutations, I compared alterations detected in blood with changes in other tissues in a patient-specific manner. Identical genomic changes detected in blood and other tissue samples in the same patient were considered germline mutations and excluded from further analysis. From a total of 749 regions detected in AK, 28 (3.7%) were excluded as germline mutations, 18 of a total 167 in NSE (10%) and 29 of a total of 161 in SE (17%). Additionally, the vast majority of blood SCNAs were random and showed no common pattern. I detected a total of 3 SCNAs in 4 patients (more than 10% of sample population). One patient had only a single SCNA in blood, and one patient had no detectable SCNAs in any of the samples analysed. These two patients were excluded from downstream analysis due to suspected technical issues (improper hybridisation), leaving data from a total of 36 individuals.

---

[11] Fios Genomics Ltd., Edinburgh, Scotland.

3.3.1.2. Characteristics of analysed samples

There were a total of 27 (75%) male and 9 (25%) female study participants. Of 27 males, 15 (55%) were immunocompetent and 12 (44%) were organ transplant recipients (OTR). Among 9 females, 2 (22%) were immunocompetent and 7 (78%) were OTR. OTR patients represented 47% of all study participants.

Somatic SCNAs detected in individual patients are summarised in Table 3.1 along with patient characteristics.

Patient ID	NSE SCNAs	SE SCNAs	AK SCNAs	**Immune Status**	**Gender**
10	2	3	10	IC	Male
45	20	0	104	IC	Male
19	0	1	0	IC	Female
11	13	2	1	IC	Male
53	0	0	3	IC	Male
29	1	3	38	IC	Male
26	4	3	4	IC	Male
47	0	2	3	IC	Male
20	35	3	4	IC	Male
52	1	2	217	IC	Female
1	8	3	22	IC	Male
41	2	1	4	IC	Male
16	0	2	11	IC	Male
46	4	6	25	IC	Male
25	0	28	4	IC	Male
3	2	2	44	IC	Male
50	3	4	2	IC	Male
24	0	1	0	OTR	Female
28	0	0	13	OTR	Male
14	4	4	1	OTR	Male
5	0	1	1	OTR	Female
33	0	4	0	OTR	Female
21	2	0	2	OTR	Female
36	1	1	1	OTR	Female
27	1	1	12	OTR	Female
31	0	0	2	OTR	Female
18	3	7	2	OTR	Male
51	4	8	66	OTR	Male
34	1	12	7	OTR	Male
9	12	2	7	OTR	Male
43	3	5	8	OTR	Male
15	11	2	91	OTR	Male
7	2	11	0	OTR	Male
44	4	5	4	OTR	Male
22	0	1	0	OTR	Male
37	4	3	6	OTR	Male
Total	148	131	719	17 IC 18 OTR	9 Female 27 Male

Table 3.1 Summary of detected somatic SCNAs and clinical characteristics of study participants.

3.3.1.3. SCNA distribution in males and females, transplant recipients and immunocompetent patients

The relative distribution of SCNAs detected in individual patients along with their baseline characteristics are shown in Figure 3.3 and Figure 3.4. If stratified by immune status, the overall number of SCNAs is relatively equal between the two groups (except for three outliers). If stratified by gender, except for those outliers, the relative contribution of individual males and females is also evenly distributed.

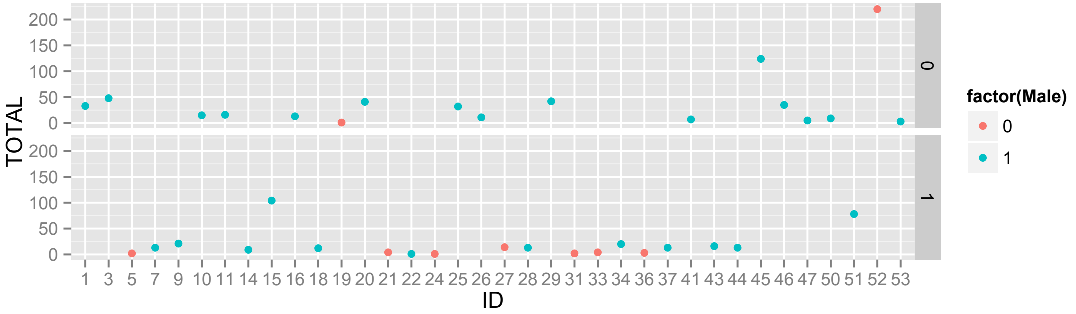


Figure 3.3 Total number of SCNAs contributed by individual patients. Pink dot - female, green dot - male. Upper panel – immunocompetent patients, lower panel – immunosuppressed patients. Two female participants were immunocompetent, one of which contributed the majority of SCNAs detected in females.

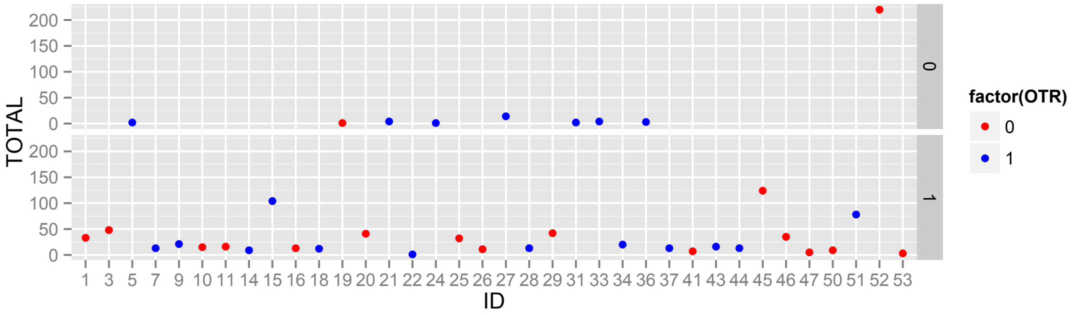


Figure 3.4 Total number of SCNAs contributed by individual patients. Red dot - OTR, blue dot - IC. Upper panel - females, lower panel - males.

The cumulative percentage of SCNAs per tissue per gender is shown in Figure 3.5. It demonstrates that for both groups, the majority of SCNA was contributed from AKs. Similar trend was observed among cumulative percentages in OTR and IC groups depicted in Figure 3.6.

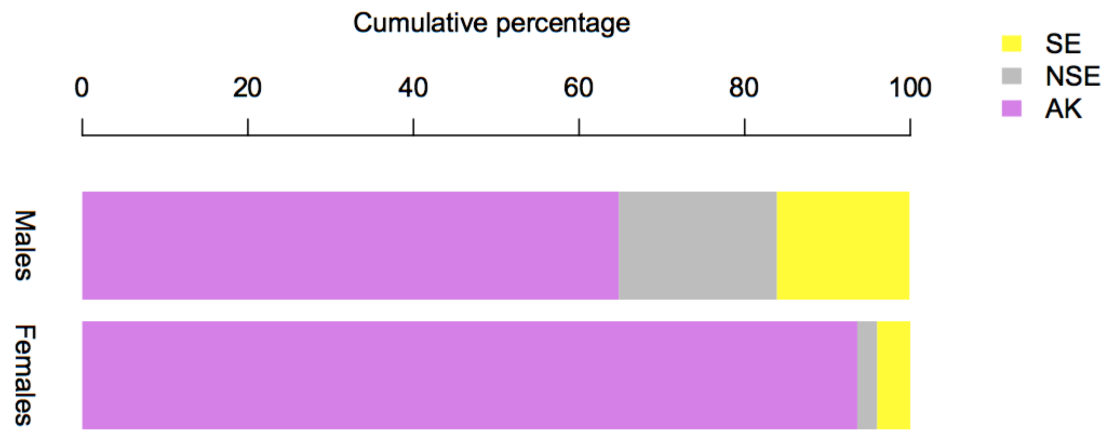


Figure 3.5 Cumulative percentage of SCNA in SE, NSE and AK in both genders.

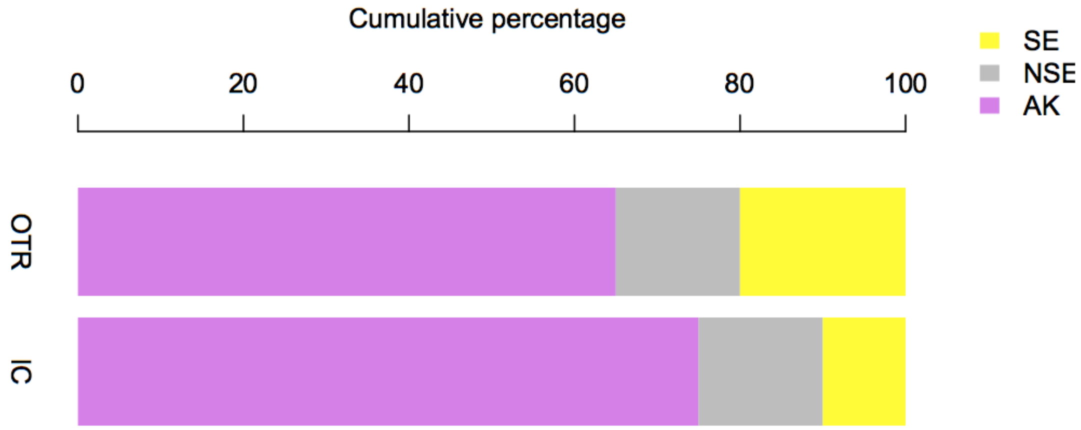


Figure 3.6 Cumulative percentage of SCNA in SE, NSE and AK in OTR and IC patients.

Summary baseline statistics for all four groups (males, females, OTR, IC) are provided in Table 3.2. Although the mean levels of SCNA fluctuate across the groups, median ranges are similar. Given the uneven distribution of samples across all groups and the uneven contributions of individual samples, median is a more accurate statistic for evaluating SCNA distribution across groups and tissue types.

Group	Statistic	AK SCNAs	NSE SCNAs	SE SCNAs	**Total**
Males	**Median**	6	3	3	15
	Mean	17.96	5.26	4.48	27.67

	Mode	4	0	2	13
	Total	484	142	121	747
	%	64.8	19	16.2	-
Females	**Median**	7	2	2	24.33
	Mean	48.29	10.6	8.68	67.98
	Mode	4	0	1	2
	Total	235	6	10	251
	%	93.6	2.4	3.99	-
OTR	**Median**	4	2	2	16
	Mean	29.18	5.65	3.7	38.53
	Mode	4	0	2	N/A
	Total	496	96	63	655
	%	75.7	14.6	9.6	-
IC	**Median**	2	2	2	13
	Mean	11.77	2.74	3.58	18.1
	Mode	0	0	1	13
	Total	223	52	68	343
	%	65	15	20	-

Table 3.2 Summary statistics of SCNA distribution per gender and immune status.

I then plotted chromosomal distribution of detected SCNAs in each sample type (SE/NSE/AK) stratified by gender and immune status (Figure 3.7). The distribution across individual chromosomes varied (males versus females, OTR versus IC), although changes on chromosome 9 dominated in all groups.

I then tested the presence of at least 1 SCNA per tissue (SE, NSE, AK) type per group (males vs. females, OTR vs. IC) with Chi square test, which showed no statistically significant difference for comparison of immunocompetent and immunosuppressed patients, but detected borderline significance for AK in comparison of males versus females, with statistically significant increase in risk difference for AK in males as shown in Table 3.3, suggesting that AK collected from male patients are more likely to contain genomic imbalances.

I finally tested this hypothesis with regression analysis, which confirmed significant differences in the total number of SCNAs by tissue type (p<0.0001) and patient ID (p<0.03). Neither immune status nor gender were significant factors for SCNA number (p value 0.47 and 0.8, respectively).

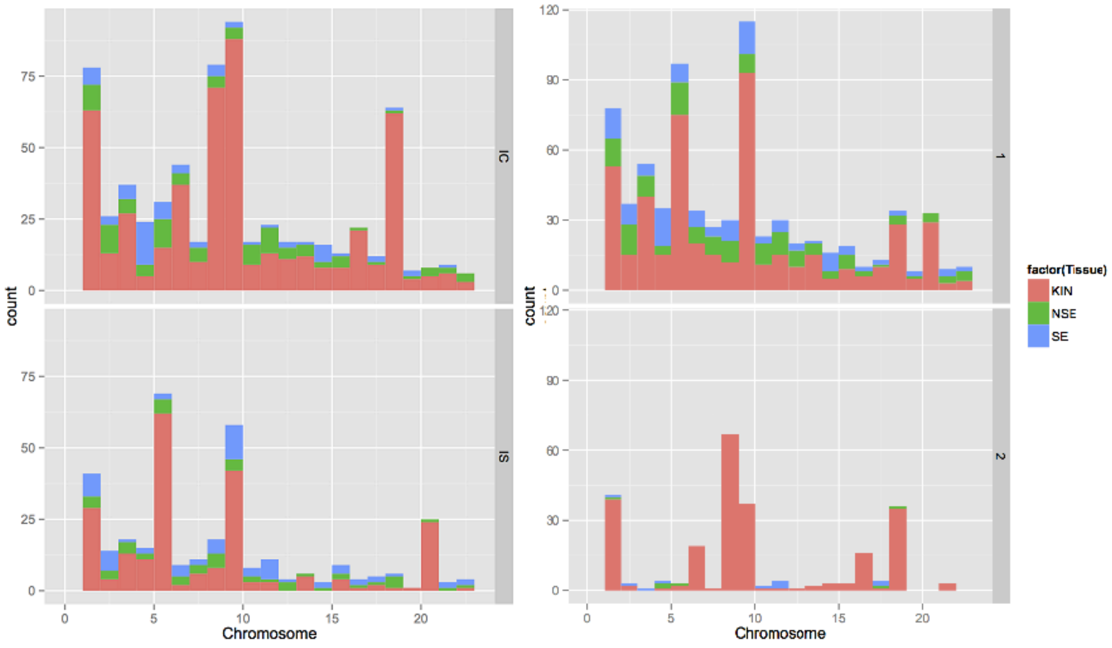


Figure 3.7 Histogram of SCNAs detected in SE, NSE and AK (KIN) stratified by patientr immune status (left) and gender (right). Upper left panel (IC) – immunocompetent, lower left panes (IS) – immunosuppressed. Upper right panel (1) - males, lower panel (2) - females. This graph depicts higher total number of SCNA regions in immunocompetent patients namely on chromosomes 1, 6, 9, 16, 18 in AK, and in males on chromosomes 1-15, 19-22.

Comparison	Tissue	Pvalue	Relative risk	Risk difference	95% CI
Males vs. Females	NSE	0.061	1.75	0.33	0.981542 to 4.194966
	SE	0.4	1.14	0.1	0.855716 to 1.985473
	AK	0.05	1.8	0.42	1.106989 to 4.956808
OTR vs. IC	NSE	0.88	0.96	0.02	0.61026 to 1.558495
	SE	0.72	0.95	0.04	0.69163 to 1.319434
	AK	0.18	0.83	0.15	0.595698 to

Table 3.3 Statistical comparison of SCNA as categorical variable between two baseline groups (gender, immune status). Risk difference (also known as absolute risk reduction) is the difference in percentage of patients with the outcome of interest. It measures the absolute size of difference between the two groups. If the confidence interval contains 0, such difference is not considered statistically significant. For AK in males, the risk of AK containing SCNA is 1.8 compared to females (80% increase), and the risk difference of 42%.

3.3.1.4. Differences between tissue types

There were no statistically significant differences between the number of SCNAs in NSE and SE (t.test, p value >0.05), yet there were differences between NSE, SE and AK (p value 0.0077, ANOVA). Differences in chromosomal distribution of SCNA across the three tissue types are shown in Figure 3.8, along with the size of individual chromosomal changes.

There was no statistically significant difference in somatic copy number alteration type (loss, gain) per tissue (linear regression, p value >0.05), but SCNA size was statistically significantly different across the tissues (p value < 0.01).

SCNA size was not different between NSE and SE (t test, p value >0.05), but ANOVA confirmed the difference in SCNA size between SE, NSE and AK (p value <0.05). Thus, SCNAs in AK are not only significantly more common, but also significantly different in size.

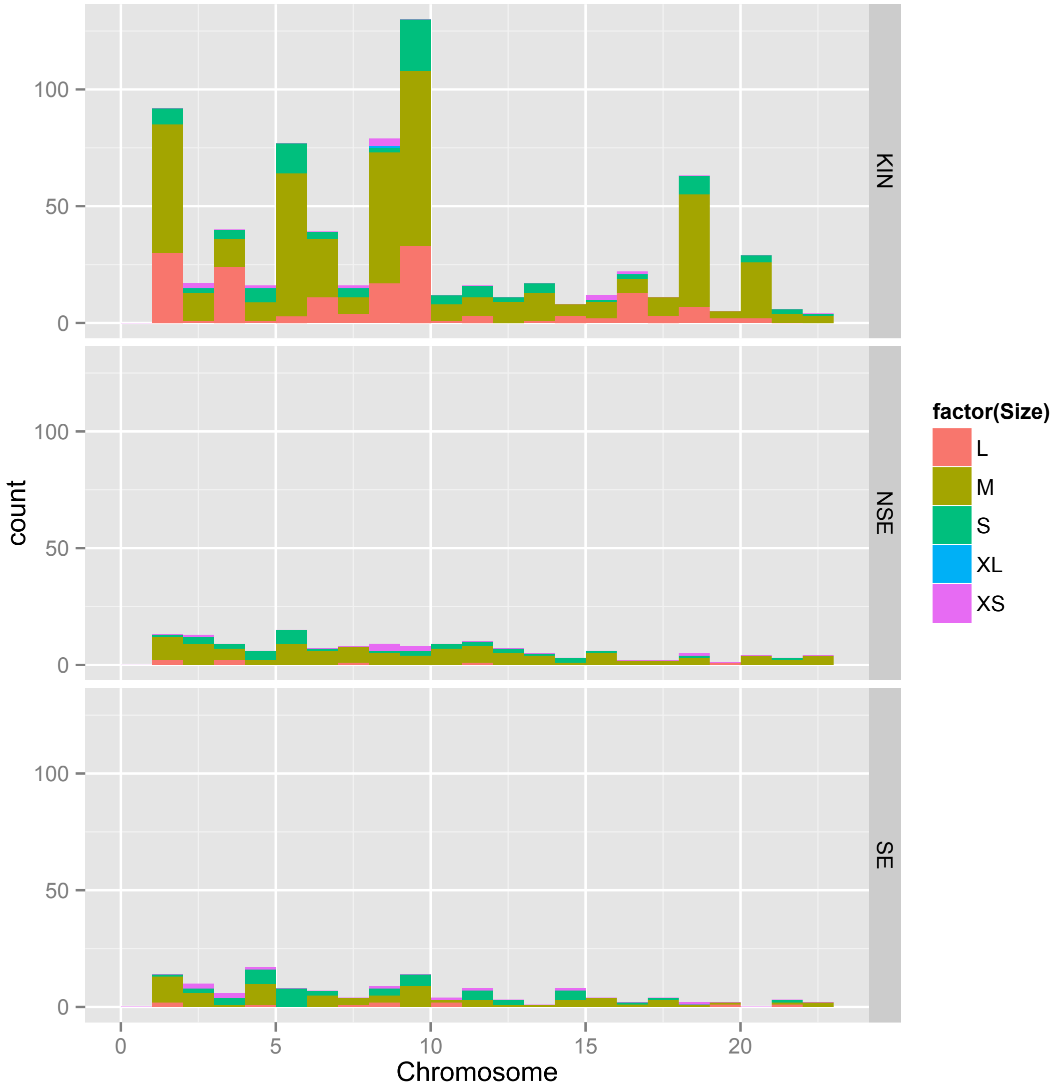


Figure 3.8 Histogram of SCNAs across chromosomes reflecting size and tissue type. KIN=AK. Large SCNAs were detected mainly in AK in chromosomes 1, 3, 5, 6, 8, 9, 16 and 18. Extra small SCNAs, on the other hand, were detected in NSE on chromosomes 2, 8, 9 and 18, and in SE on chromosomes 2, 3, 4, 8, 11, 14 and 16.

Additionally, copy-number (CN) significantly differed across chromosomes and between patients, and was size-dependent (linear regression). The majority of changes detected were monoallelic gains (723/1003, 72%), followed by monoallelic losses (221/1003, 22%), biallelic losses (42/1003, 4.2%) and 17 double gains (1.6%). Figure 3.9 shows that the biallelic losses or double gains were predominantly small in size, while the biggest changes in terms of size were monoallelic gains.

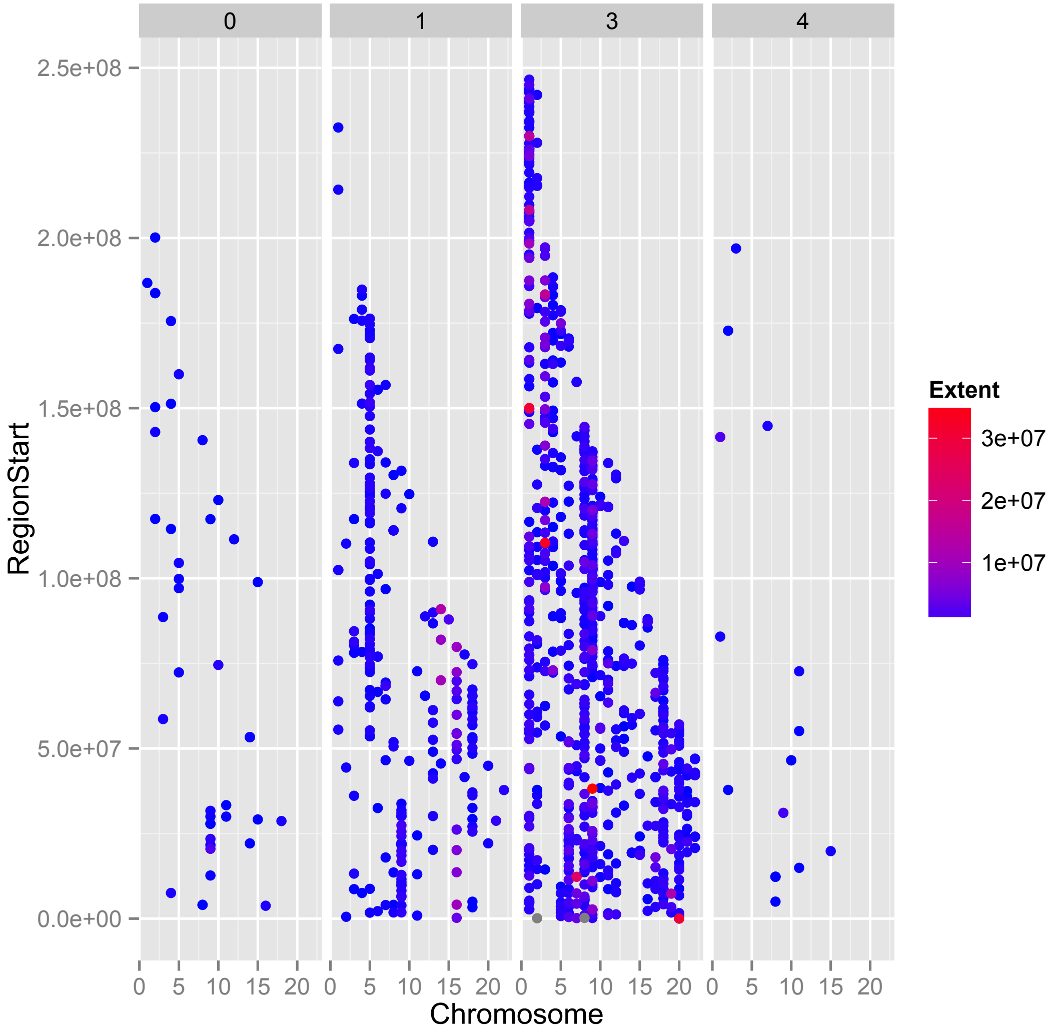


Figure 3.9 Distribution of SCNAs per chromosome divided into groups according to SCNA characteristic and size in base pairs. RegionStart shows genomic coordinate of SCNA on a chromosome, groups reflect change in genomic material (0=biallelic loss, 1=monoallellic loss, 3=monoallelic gain, 4=biallelic gain). Extent shows a gradient of increasing size of SCNA in base pairs. Groups 0 and 4 contain mainly small size SCNAs, while group 3 contains several large gains on chromosomes 1, 3, 7, 9 and 20 represented by red dots.

I summarised the distribution of SCNA size per chromosomal change in Table 3.4, and observed that 17 of 27 XS changes (63%) are biallelic loss or double gain SCNAs. In comparison with the number of XS monoallelic lesions, this is a highly significant finding (p value<0.0001, Fisher's exact test). The same was true for small (S) lesions as well.

Size	CN0	CN1	CN3	CN4
XS	9	6	4	8
S	21	49	97	3
M	11	144	468	4
L	1	22	153	2

XL	0	0	1	0

Table 3.4 Summary of the number of SCNA based on size and the change of genomic material. CN0=biallelic loss, CN 1=monoallellic loss, CN3=monoallelic gain, CN4=biallelic gain.

I then plotted SCNA size distribution as bin value based on XS to XL size classificatioin (Figure 3.10) rather than individual values shown in Figure 3.9, and observed clustering of large SCNAs among monoallelic gains. Those represented a total of 21% of all CN3, while only 10% of CN4, 9% of CN1 SCNAs, and 2% of CN0, which was also a statistically significant finding (p value<0.05). Moreover, I observed clustering of changes across individual chromosomes. Specifically, there seemed to be a concentration of M size monoallelic losses across chromosome 5, p-arm of chromosome 9, 13 and 18. There were also clusters of monoallelic gains on chromosome 1, 3, 6, 8, 9, 18 and 20. Distribution of M size SCNAs was equal across groups, while all other size categories except for 1 XL SNP showed significant association with CN (small were predominantly biallelic losses or double gains, large were monoallelic SCNAs).

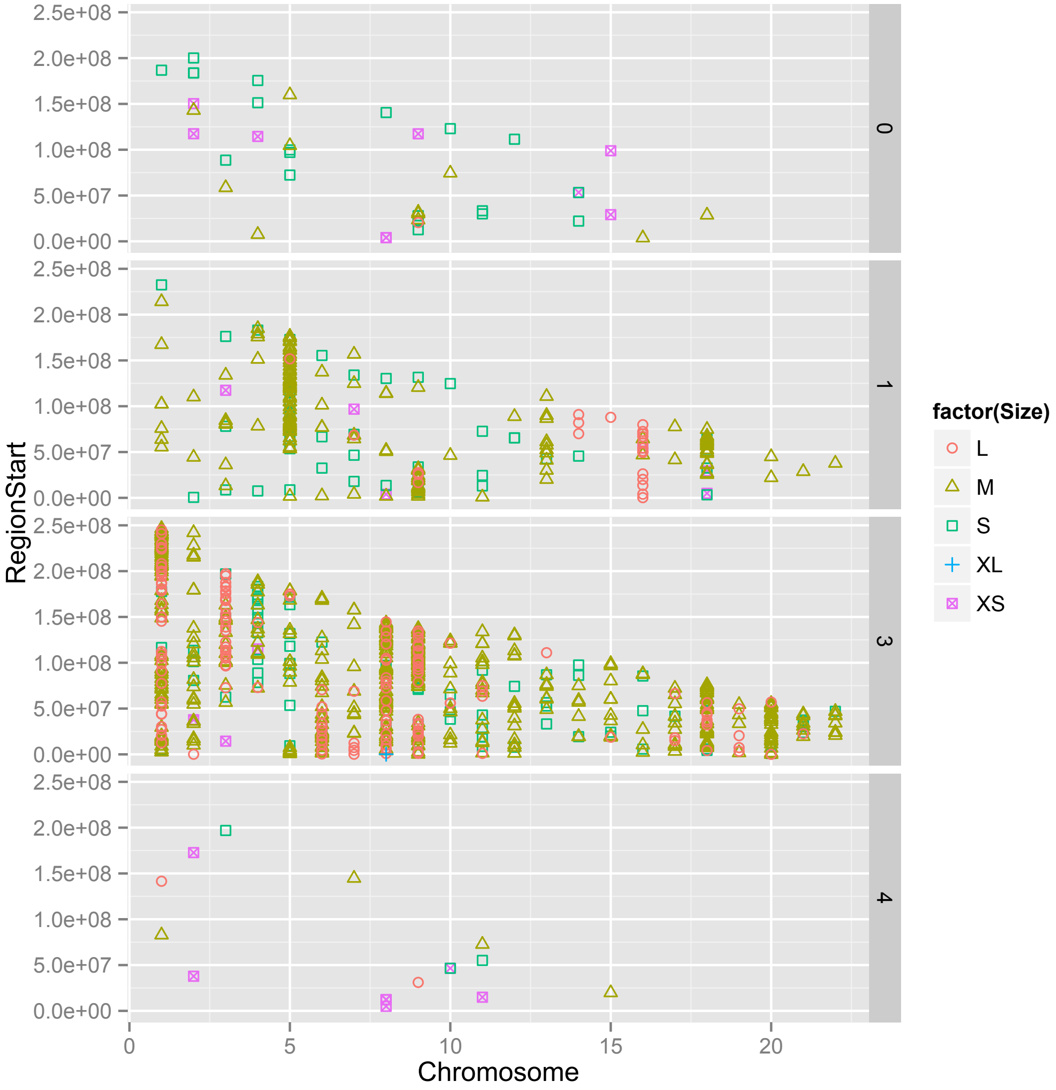


Figure 3.10 SCNA distribution across CN groups plotted with SCNA size as bin factor. 0=biallelic loss, 1=monoallellic loss, 3=monoallelic gain, 4=biallelic gain. This figure shows XS aberrations concentrated in groups 0 and 4, while several M-sized monoallelic losses occurred on chromosome 5, 9 and 16. Chromosome 1 contained mainly M- and L-sized monoallelic gains, and similar trend was apparent on chromosomes 3, 6, 8, 9 and 16 in group 3.

The chromosomal distribution of CN and SCNA size stratified by tissue type is shown in Figure 3.11. It shows that clusters of CN1 on chromosome 5, p-arm of chromosome 9, 13 and 18 were predominantly derived from AK lesions, as were most of the conspicuous gains.

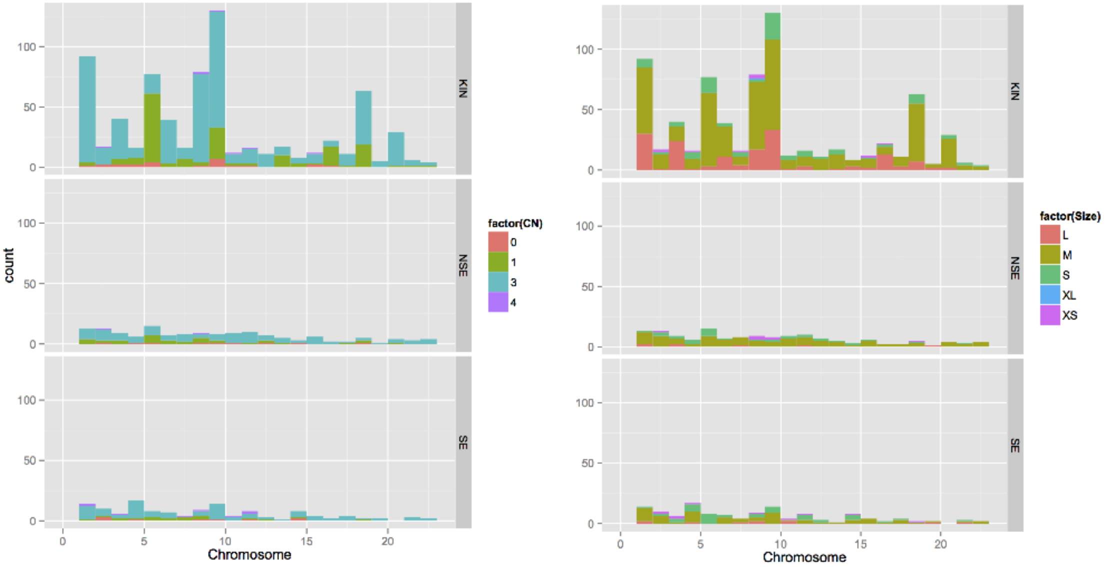


Figure 3.11 Histogram of copy-number change and SCNA size chromosomal distribution across all tissue types. Copy-number changes in AK (top left panel) show both prominent monoallelic gains and losses on chromosomes 5, 7, 9, 13, 16 and 18. In NSE, similar co-occurrence of loss and gain of a genomic material is found on chromosomes 1, 3, 5, 6, 8 and 18. In SE, such finding can be observed on chromosomes 3, 5-8. Right panel corresponds to Figure 3.8 and is included for reference.

A more detailed look at distribution of SCNAs per tissue with size and CN factored in is shown in Figure 3.12. While it confirmed previous observations, it has also demonstrated that there is relatively little overlap between SE and NSE SNPs. While some were clearly identical, the majority was tissue-specfic.

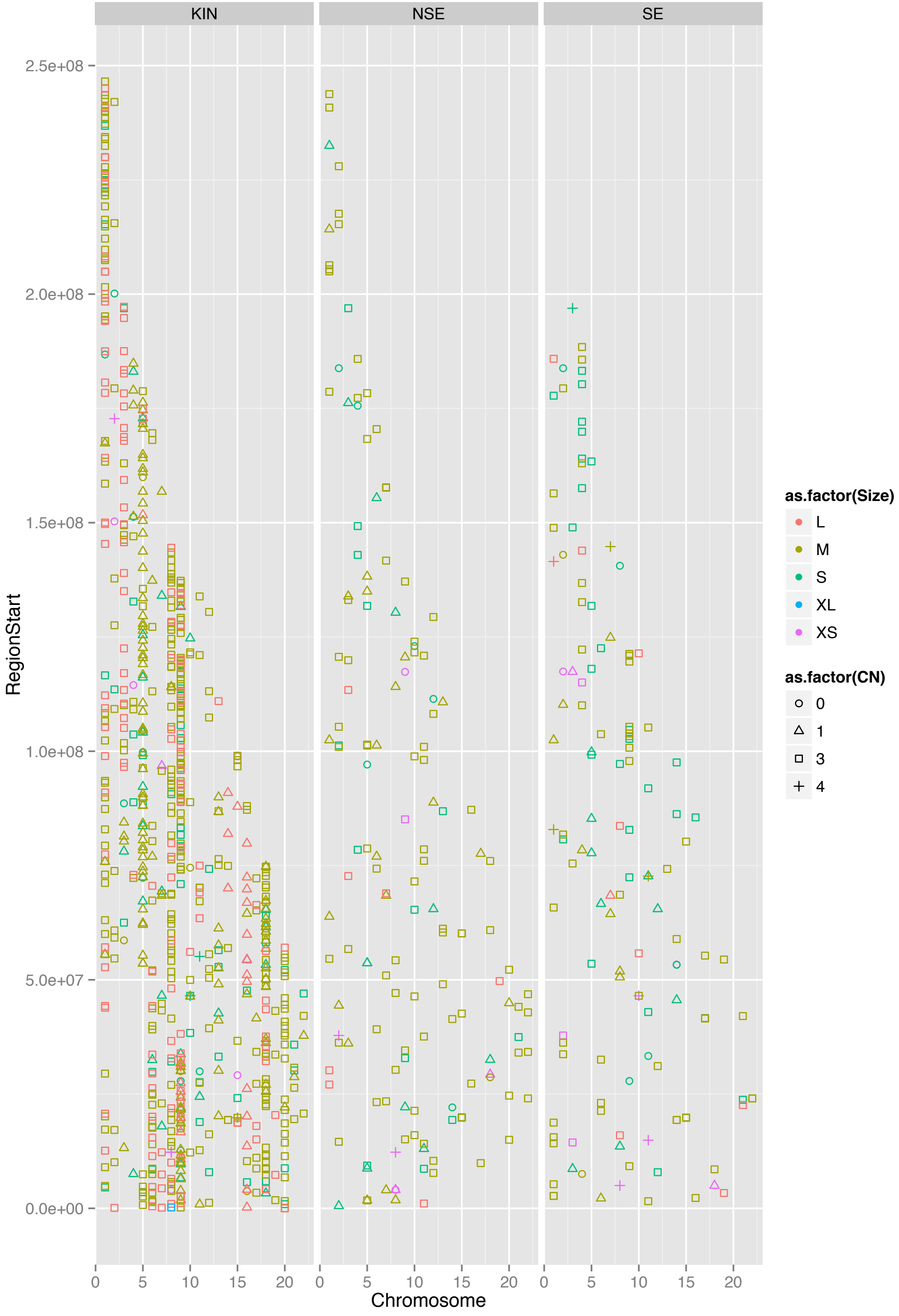


Figure 3.12 SCNA distribution across tissue types with chromosomal location. Substantial increase in both the number and the size of genomic aberrations is apparent in AK (KIN) samples, many of which affect entire chromosomes (e.g. chromosome 1, 5, 8, 9,18, 20). Normal skin samples, on the other hand, contain predominantly small and random aberrations.

I then identified shared, identical SCNA across NSE and SE samples. There were a total of 11 (3.9%) shared SCNAs listed in Table 3.5. Three of those shared SCNAs occurred in 1 patient (OTR male), two in another patient (OTR male), and 2 of those 11 changes were shared by 2 patients. 1 SCNA showed monoalelic gain in SE and double gain in NSE of the same patient. Given that the majority of the SCNAs shared between NSE and SE occurred in the same patient (8/11, 72%), this further confirmed absence of shared genomic traits in those samples, and the impact of individual patients as the key contributing factor to the character of SCNAs in each tissue.

SNP	Chr	Start	End	ID	Gene	CN	Tissue
SNP_A-4216701	1	102430762	102622376	ID15	OLFM3	1	NSE/SE
SNP_A-1900201	2	183794033	183819678	ID1	NCKAP1	0	NSE/SE
SNP_A-2218448	2	36244590	36422518	ID15	-	3	NSE/SE
SNP_A-4194705	2	37814187	37814499	ID51	-	4	NSE
SNP_A-4194705	2	37814187	37814499	ID51	-	3	SE
SNP_A-2056690	3	196918723	196959312	ID14	DLG1	3	NSE
SNP_A-2056690	3	196918723	196959312	ID16	DLG1	4	SE
SNP_A-1843546	5	131806615	131844802	ID50	C5orf56, IRF1	3	NSE/SE
SNP_A-2225884	7	68428002	69429303	ID43	AUTS2	1	NSE/SE
SNP_A-2162694	12	65484396	65540472	ID37	WIF1	1	NSE/SE
SNP_A-4233185	14	19336854	19424016	ID41	OR11H12, LOC642426	3	NSE
SNP_A-4233185	14	19336854	19456378	ID10	OR11H12, LOC642426	3	SE
SNP_A-2228752	15	19821421	19943185	ID14	-	3	NSE/SE
SNP_A-2253908	22	24049288	24241071	ID14	GUSBP11, ZNF70, VPREB3, C22ORF15, CHCHD10, MMP11, SMARCB1, DERL3, SLC2A11, MIF	3	NSE/SE

Table 3.5 Summary of SCNAs shared between NSE and SE.

3.3.1.5. Differences between males and females

While the presence of SCNAs did not differ between the two groups, I tested the hypothesis that there may be differences in the SCNA size and copy number. However, linear regression - factoring in patient ID - demonstrated no association between gender and SCNA size or CN. Although Figure 3.13 shows rather intriguing differences between the two groups in chromosomal distribution, Figure 3.14 and Figure 3.15 display the number of SCNA regions per tissue type per person, suggesting the majority of SCNAs in group 2 (females) are derived from a single patient, reflecting the low number of female participants in the study. A detailed look at SCNAs contributed by remaining female participants is shown in Figure 3.16.

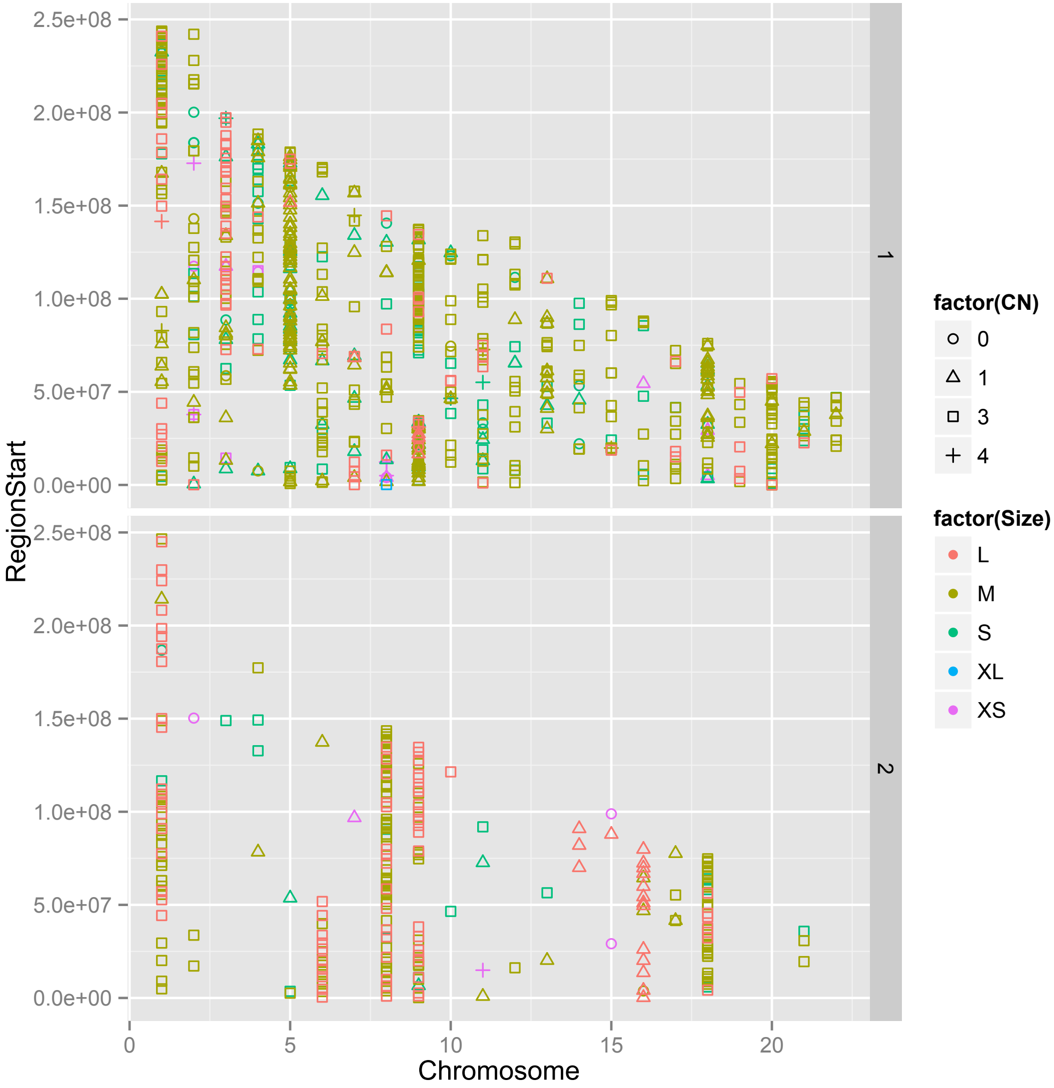


Figure 3.13 SCNA region size and CN distribution per gender. 1-males, 2-females. 0=biallelic loss, 1=monoallellic loss, 3=monoallelic gain, 4=biallelic gain. This graph shows prominent differences in both chromosomal location, size and CN character of SCNAs between males and females: many chromosomes in female patietns are not at all or minimally affected by SCNAs,

with a concentration of monoallelic losses on chromosome 16 and clusters of gains on chromosomes 1, 6, 8, 9, 16 and 21. Male participants demonstrate greater heterogeneity in both the size and character of genomic aberrations on individual chromosomes.

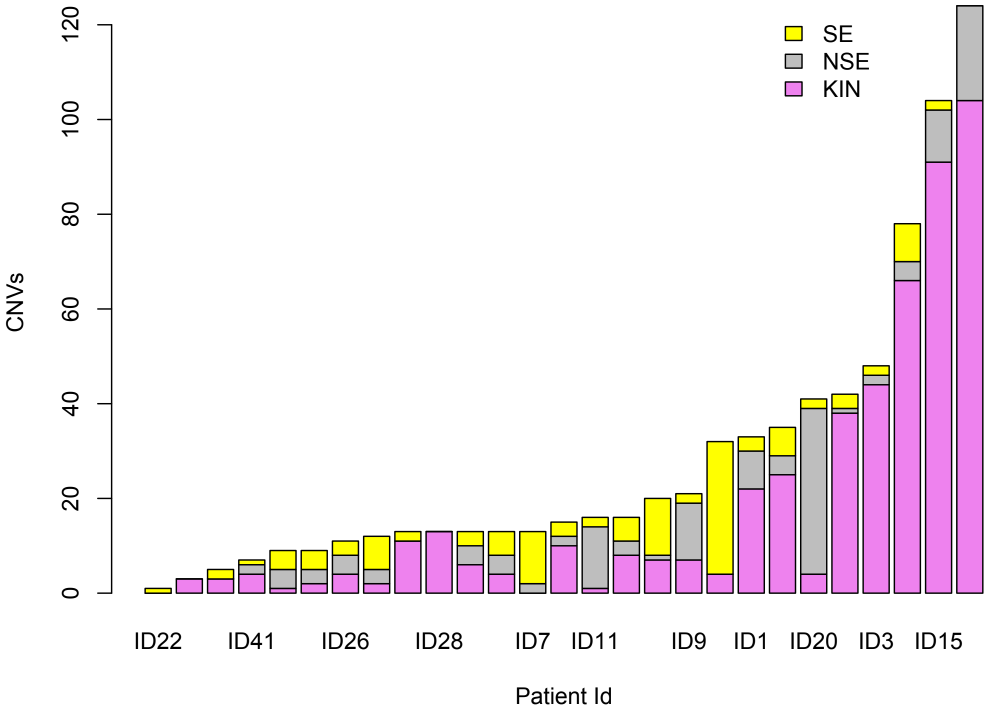


Figure 3.14 SCNA distribution per tissue type in individual male patients. Most SCNA events were contributed by AK (KIN) samples, with increased proportion of NSE SCNAs in patients 11, 9 and 20.

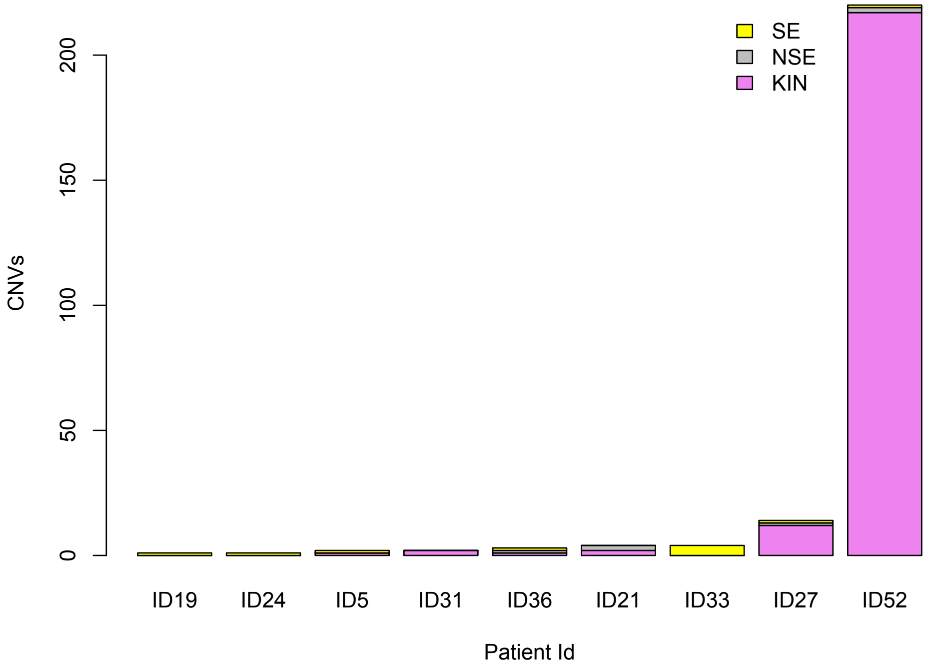


Figure 3.15 SCNA distribution per individual female participants. Patient 52 contributed the vast majority of SCNAs in this group.

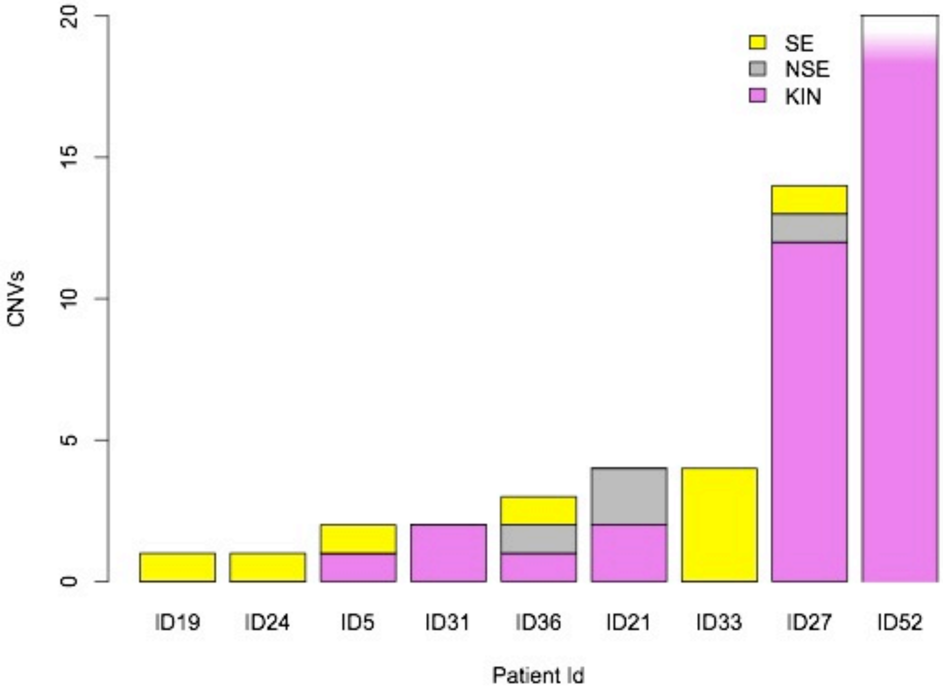


Figure 3.16 SCNAs in female participants, patient 52 not included in the display range (column with gradient). Three female patients had detectable SCNAs only in their SE skin.

I wanted to verify the absence of overlapping differences between the two groups in terms of SCNA distribution in AK lesions. As seen in Figure 3.17, while there were some shared traits, namely on chromosomes 1, 8 and 9, there seemed to be no obvious pattern among the lesions, other than as a reflection of the fact that there were more male participants and thus more SCNAs in this group.

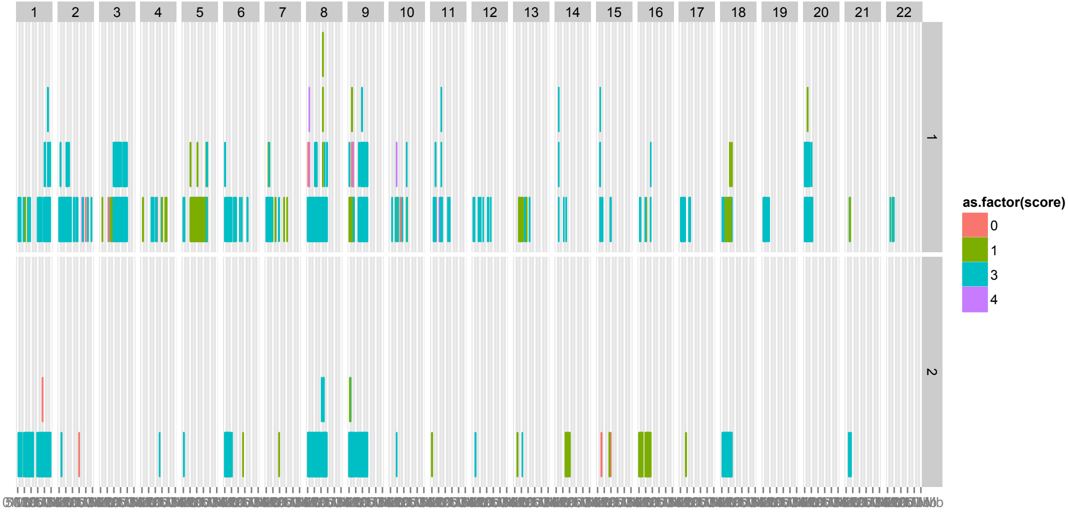


Figure 3.17 SCNA distribution across chromosomes in AK. Score=CN, upper panel (1) - males, lower panel (2) - females. The size of coloured block corresponds to genomic coordinates of each detected SCNA. Overlapping SCNAs are plotted above one another. This graph shows that the highest number of overlapping SCNAs was four on chromosome 8 in male patients, in addition to differences in copy-number changes of overlapping regions: for example, two overlapping regions in male patients on chromosome 20 are monoallelic gains (blue), and 1 smaller overlapping SCNA region is a monoallelic loss (green).

A more detailed look at SCNAs across chromosome 1 in AK is shown in Figure 3.18. While both the location and distribution of SCNAs are vastly different between males and females based on this plot, as shown in Figure 3.19, the majority of those SCNAs are derived from 1 male and 1 female patient, further stressing the key role of individual genomic background.

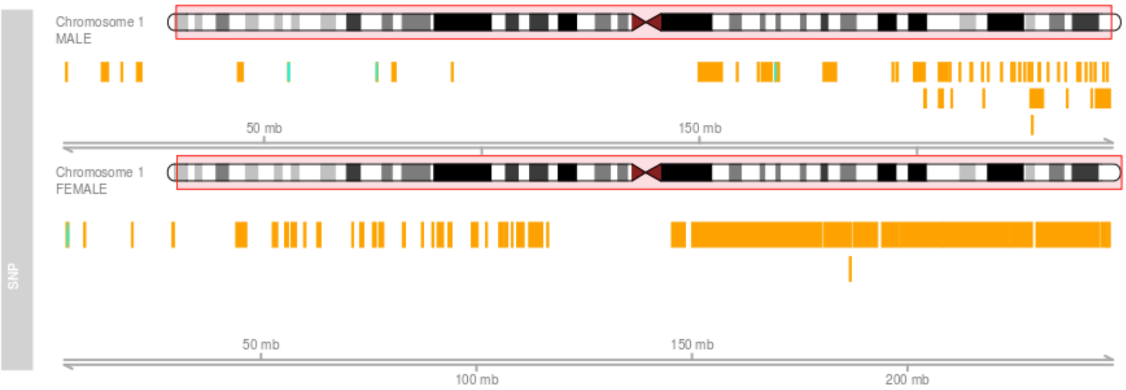


Figure 3.18. SCNAs detected on chromosome 1 in AK. Geen=loss, orange=gain. This detailed look depicts many short SCNAs on the p-arm in female patients, while such cumulation of short SCNAs occurs on the q-arm in male patients.

Figure 3.19 SCNAs distributed per chromosome per gender in individual patients. Red-males, green-females. Each line in the matrix corresponds to patient ID (y axis), and SCNAs are plotted based on their chromosomal coordinates (x axis). Each column corresponds to an individual chromosome. This graph illustrates how each individual patient contributed to the overall changes detected with the SNP arrays, and that the majority of chromosomal aberrations are in fact short (focal), rather than arm-long. For example, it shows large regions of SCNA in patient 52 across 7 chromosomes, while many shorter SCNAs in patient 45 across 21 chromosomes.

3.3.1.6. Differences between organ transplant recipients

I then proceeded to find out if there were differences between organ transplant recipients (OTR) and immunocompetent (IC) patients. Given that the numbers of participants in each group were almost equal (19 and 17, respectively), such differences would be more representative of trends in these two groups.

There were no significant differences in the number of SCNAs per patient in SE, NSE or AK between the two groups (treating SCNA as a continuous variable, T-test, p =0.9, 0.24 and 0.2, respectively).

Linear regression analysis has shown that while the size of the SCNA was independent of immune status, CN character (gain or loss) was largely associated with immune status (p value < 0.0001). A cumulative percentage plot shown in Figure 3.20. While over 80% of SCNAs among IC patients were gains, over 40% among OTR (IS) patients were losses. Moreover, both biallelic losses and double gains were more common among OTR patients, but this difference was not significant (Chi square, p value =0.8). The CN1 (hererozygous deletion) was statistically significantly more common among SCNAs detected in OTR patients (p value <0.0001, Chi sq test).

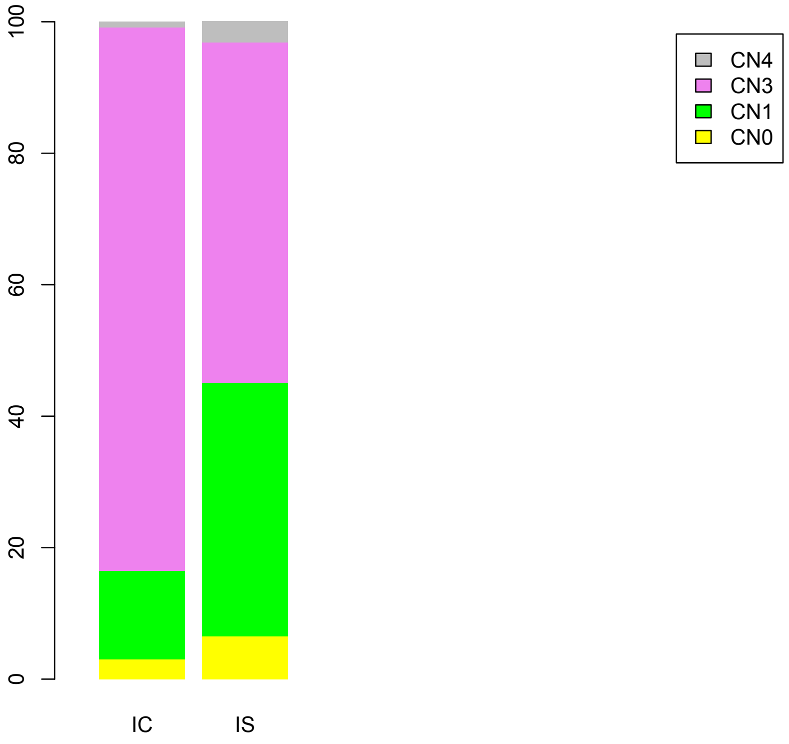


Figure 3.20 Cumulative percentage of copy number segregated by immune status. IC=immunocompetent, IS=immunosuppresed patients. 0=biallelic loss, 1=monoallellic loss, 3=monoallelic gain, 4=biallelic gain. Monoallelic losses and biallelic gains and losses were more

common among immunosuppressed patients. Generally, loss of genomic material represented almost 50% of detected chromosomal changes in this group, compared with under 20% among immunocompetent patients.

3.3.1.7. Contribution of individual patients

The contribution of individual patients towards the total number of genetic changes in each tissue type is shown in Figure 3.21. The uneven distribution of individual SCNA changes suggests these somatic mutations are predominantly stochastic, and that the extent of genomic changes is largely patient driven.

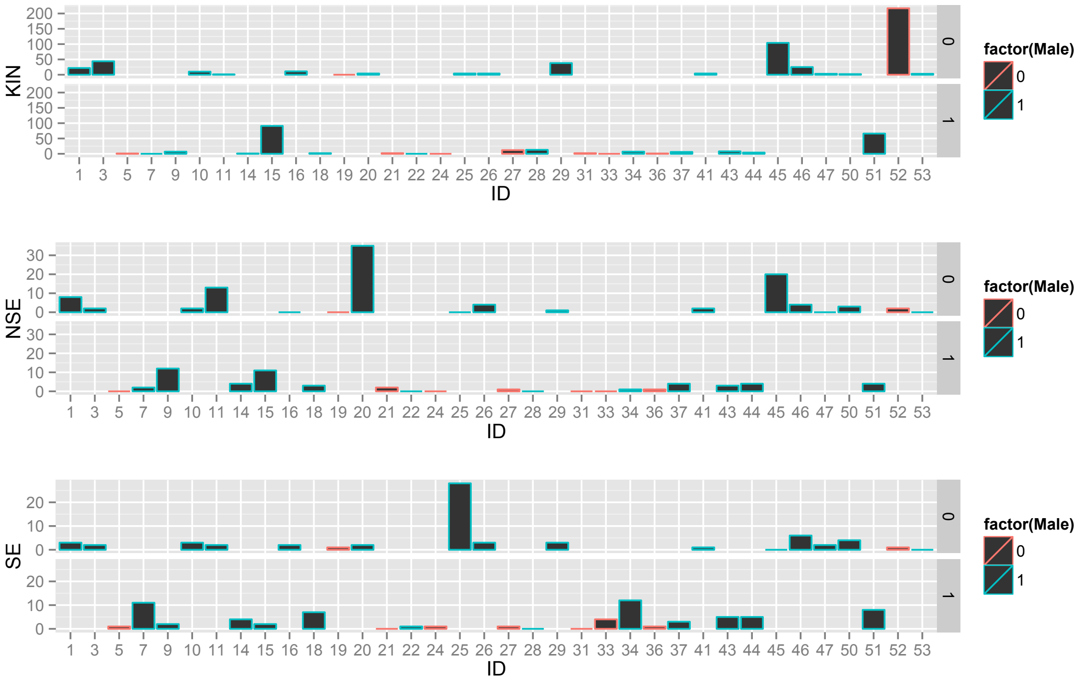


Figure 3.21 Total number of SCNA contributed by individual patients. Pink outer line - female, green outer line - male. Upper panel (0) - Immunocompetent, lower panel (1) – transplant patients. Patients 1, 3, 29, 45, 46 and 52 contributed the vast majority SCNAs in AK (KIN) among immunocompetent individuals, while patients 1, 11, 20 and 45 contributed the majority of SCNAs in NSE in this group. Patients 25, 46 and 50 contributed the majority of SCNAs in SE. The distribution across transplant patients is considerably more even compared with immunocompetent individuals.

The presence of at least 1 SCNA detected in each chromosome per tissue type is shown in Figure 3.22. It demonstrates inter-patient heterogeneity in SCNA and relative absence of shared patterns among the patients.

Linear regression analysis has confirmed that SCNA size, CN and chromosomal distribution of those regions are associated with patient ID. CN distribution is shown in Figure 3.23, clearly depicting major clustering of heterozygous gains in patient 45, 20, 1 and 25, and extensive heterozygous gains and losses in patient 52. Hererozygous losses were predominant in patients 37, 43 and 9. Patients 15 and 29 each had a large SCNA in extent. In order to achieve greater data spread, after subtracting those two biggest SCNAs and plotting SCNA size and CN per patient, it emerged that while in certain patients lesions clustered along chromosomal locations in terms of both size and CN (patient 15, 51), in many there were no patterns (Figure 3.24).

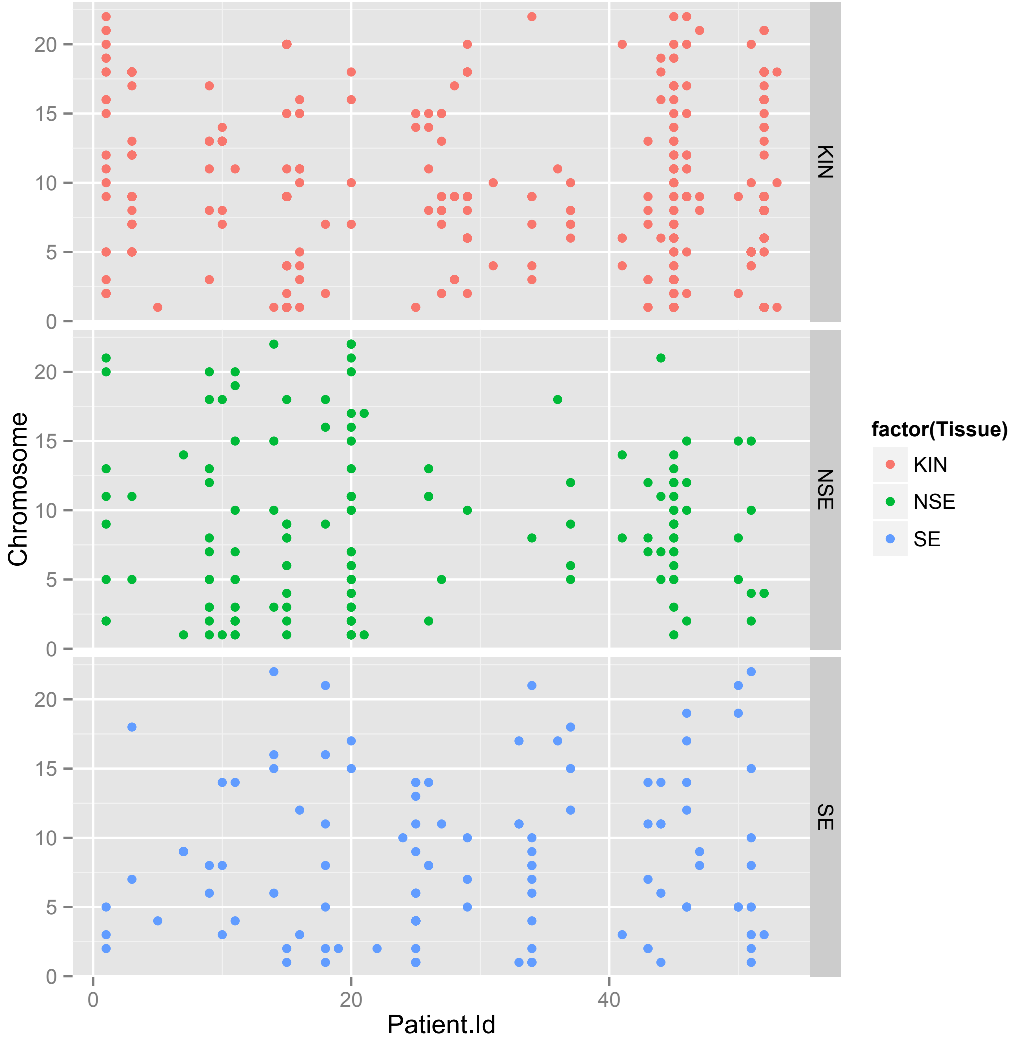


Figure 3.22 Chromosomes showing at least 1 SCNA per lesion in individual patients. Some chromosomes were consistently affected by genomic aberration in individual patients (for example, NSE and AK[KIN] overlap in patient 45), while some were random (for example, patient 41).

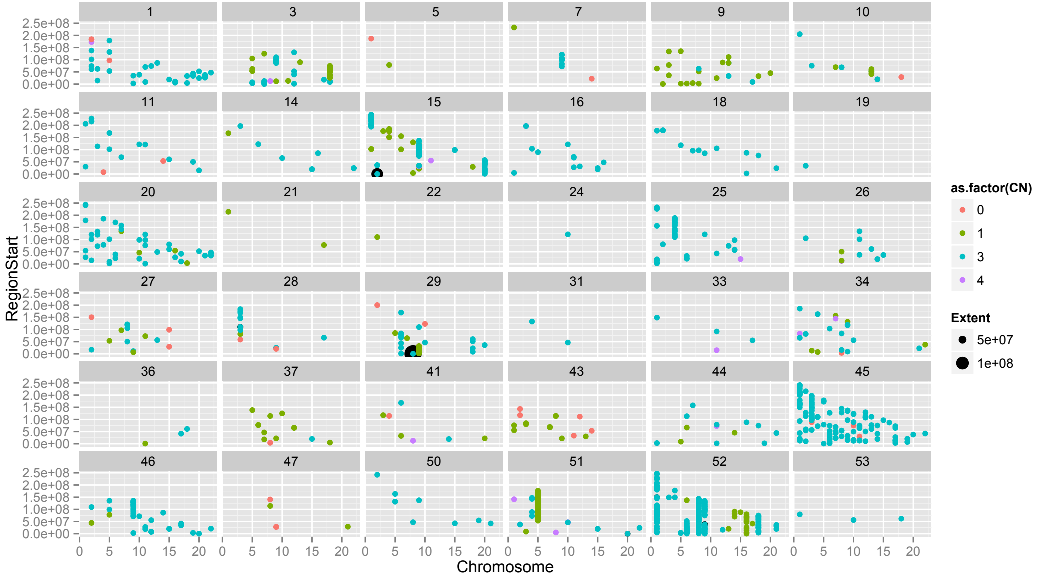


Figure 3.23 Distribution of copy number per chromosome per patient. Each box corresponds to an individual patient. 0=biallelic loss, 1=monoallellic loss, 3=monoallelic gain, 4=biallelic gain. Y axis of each box=chromosomal location of SCNA. This graph show key differences in the number and character of SCNA across individual patients. Patient 9's SCNA were almost all monoallelic losses, while patient 45's SCNAs were mostly monoallelic gains. Patient 43's SCNAs were only mono- or biallelic losses, while patient 50's SCNAs were only monoallelic gains.

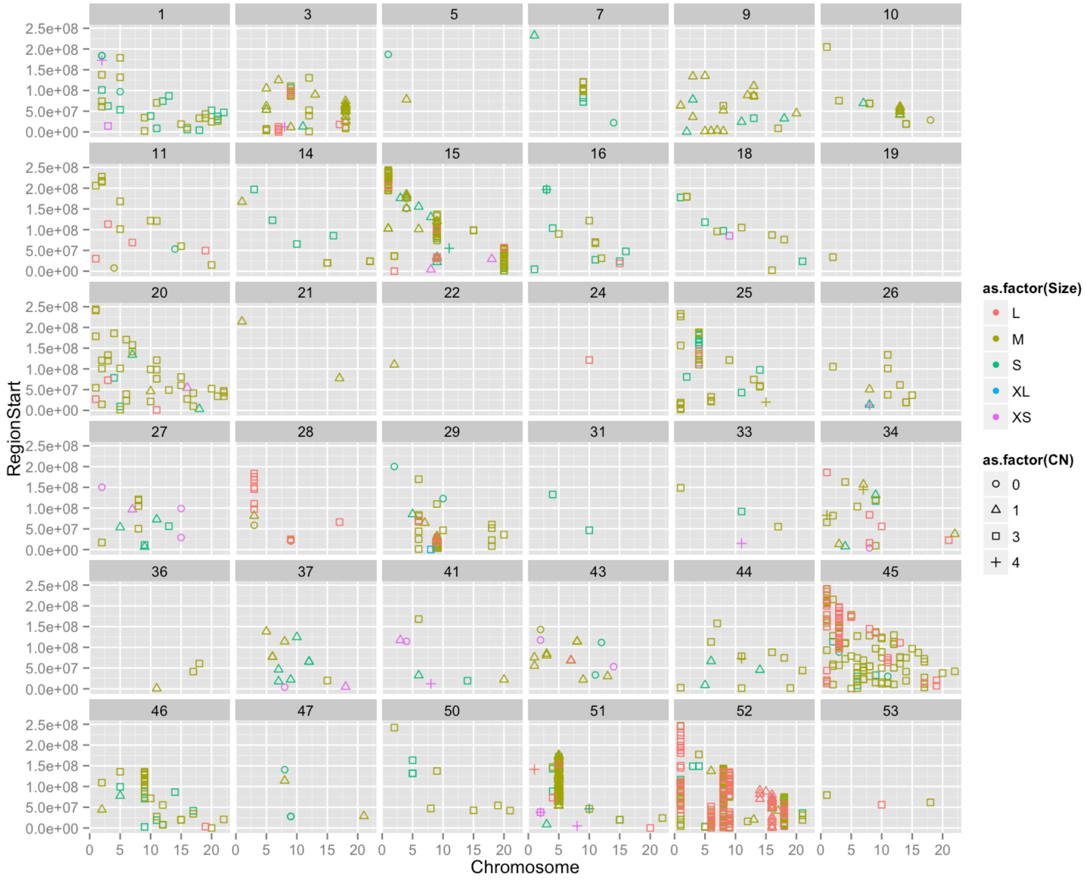


Figure 3.24 SCNA distribution per chromosome in each patient showing both CN and SCNA size. Each box corresponds to an individual patient. 0=biallelic loss, 1=monoallelic loss, 3=monoallelic gain, 4=biallelic gain. Y axis of each box=chromosomal location of SCNA. This graph shows more detailed differences in the size of SCNAs across all patients.

3.3.1.8. Multiple overlapping SCNAs in SE, NSE and AK

I proceeded to find overlapping SCNAs (not necessarily identical in their start or end, but having an overlapping segment of the SCNA) that appeared in at least 3 patients in each tissue type (8% or more).

Based on Figure 3.25, such changes occurred on chromosome 15 and 17 in SE, chromosome 15 in NSE, and chromosome 1, 6, 8, 9, 11, 14, and 15 in AK. Changes are listed in Table 3.6. If there were differences in size in the overlapping segment, only the size of the smallest segment is listed. If the overlapping change differed in loss/gain, the aberration was listed only if the shared genomic material loss or gain was shared by three or more patients.

A heterozygous loss in a single region (15.q11.1) was shared by NSE and SE, as it was present in 2/3 patients that shared the lesion in both tissues. Many of the observed overlaps were due to one or two patients accumulating aberrations spanning large chromosomal regions, rather than due to specific small changes. Example of this pattern is shown on chromosome 9, where many of SCNAs accumulated, but most differed in their CN, namely on p arm (Figure 3.26 and Figure 3.27).

Tissue	Chromosome	Cytoband	Size	CN	Gene	Exp	Id
SE	15	15q11.1	M	3	-		14,51,37
	17	17q21.31	S	3	-		20,36,46
NSE	15	15q11.1	S	3	-		14,51,46
AK	1	1q42.12	M	3	LIN9	NS	45,25,52
	1	1q43	S	3	-		45,25,52,15
	6	6p22.3	M	3	-		29,45,52
	8	8q12.3	M	3	NKAIN3	NP	9,29,52
	8	8q13.2	M	3	PREX2	NP	10,29,52
	8	8q22.3	M	3	-		27,29,52
	8	8q23.3	M	1	MED30	NS	37,43,47
	8	8q24.11	M	3	-		27,29,52
	8	8q24.12	M	3	COL14A1	NS	27,29,52
	8	8q24.21	M	3	POUF1B	NP	45,29,52
	9	9p21.2	S	3	-		46,1,52
	9	9p21.33	M	3	GAS1, C9ORF170	**UP** NP	46,15,52
	9	9p22.2	M	3	SYK	NS	15,3,52
	9	9p22.32	S	3	-		15,46,52,45
	9	9p34.3	L	3	NANS, TRIM14, CORO2A, TBC1D2, GABBR2	NS **DOWN** NS **DOWN** **NP**	15,52,3
	9	9q33.1	S	3	ASTN2	NS	15,46,52
	9	9q31.2	M	3	-		3,52,29
	9	9q34.11	M	3	ASS1, FUBP3	NS NS	15,46,52
	9	9q33.1	M	3	ASTN2	NS	46,15,52
	9	9q34.13	M	3	ADAMTS13, C9ORF96, SURF2, GTF3C5, RPLA7, GFI1B, CEL, C9ORF9, CACFD1, TSC1, RALGDS, OBP2B, SLC2A6, MED22,	NP NP NP NP NP NP NP NS NP NP NS NP NP NP	15,46,52

				SURF4, REXO4, GBGT1, SURF1	NP NS NP NP	
9	9q34.2	M	3	VAV2, BRD3, WDR5	NS NP NS	15,45,52
9	9q34.2	S	3	VAV2, BRD3, WDR5	NS NP NS	46,52,45
9	9q34.2	M	3	RXRA	**UP**	50,52,45
11	11q13.3	M	3	PPFIA1, CTTN	**DOWN** **DOWN**	1,16,45
14	14q11.2	M	3	OR11H12	NP	10,45,26
15	15q11.1	M	3,4	-		16,46,25

Table 3.6 SCNAs shared by at least 8% of patients in given tissue type. Exp=expression fold-change. UP=significantly upregulated in AK compared to SE, DOWN=significantly down-regulated, NS=statistically non-significant difference, NP=not present on expression array.

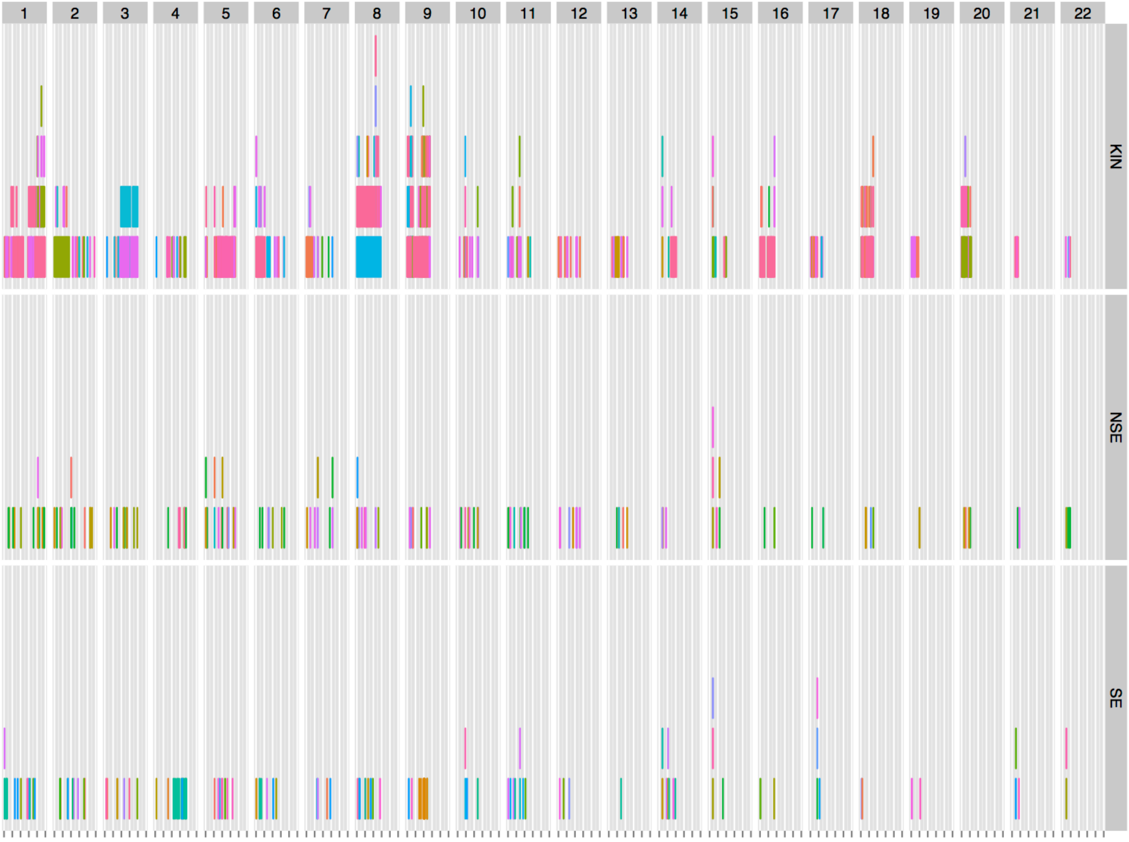


Figure 3.25 Cumulative SCNAs per chromosome in tissues coloured by individual patient ID. Each column corresponds to a chromosome, each color shade represents an individual patient

(see Figure 3.27 for legend of patient ID). Size of the SCNA corresponds to size in base-pairs. The largest SCNAs were detected in AK (KIN), and many were contributed by a single patient.

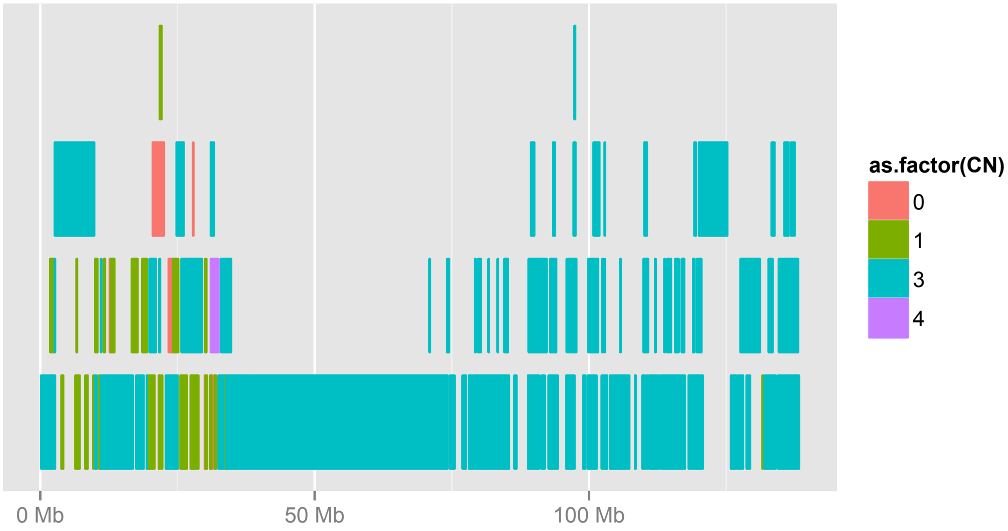


Figure 3.26 SCNA on chromosome 9 showing cumulation of discrete CN changes on p arm. P arm=0 to 50 Mb. 0=biallelic loss, 1=monoallellic loss, 3=monoallelic gain, 4=biallelic gain. Each SCNA corresponds to its genomic coordinates. While all CN types are detected on the p-arm, the q-arm (from 50 Mb on) shows almost exclusively monoallelic gains.

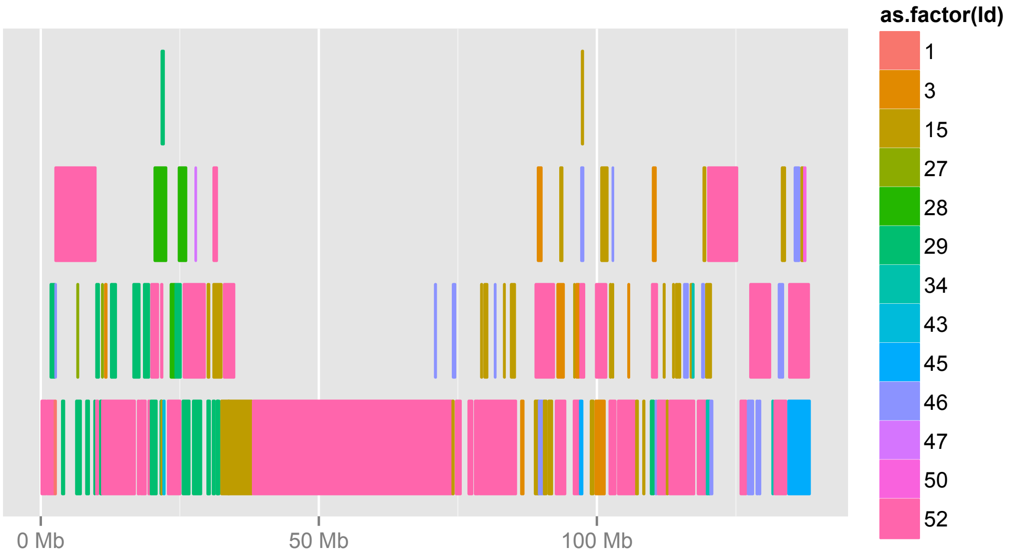


Figure 3.27 CN on chromosome 9 painted by patient ID. Patient 52 has contributed to the majority of shared changes due to having long ranges of chromosomal aberrations.

3.3.2. Transcription changes in normal skin and actinic keratoses

3.3.2.1. Experimental design and comparison strategy

Expression array profiling was generated from samples obtained from a total of 44 patients. Of these, 20 patients had at least one of their lesions profiled by SNP array as detailed in the previous section.

A total of 44 AKs, 20 of which were collected from unique patients, passed the quality control. 16 NSE samples and 20 SE samples passed the quality control and were included in the study.

Because SE and NSE skin are considered distinct tissue categories in terms of showing intrinsic genomic changes, the comparison with AK was conducted separately for each type of skin sample. Due to the relatively low number of paired samples (20), there was insufficient power to perform analyses on an individual patient basis.

3.3.2.2. Transcription profiling and statistical analysis of microarray expression data

Transcriptional profiling of NSE, SE and AK samples was generated by hybridization of 5 micrograms of amplified cDNA synthesised with NuGen WT-ovation Pico RNA Amplification System to Affymetrix HG U133 Plus 2.0 microarrays (Affymetrix, USA). This chip comprises 54,675 probes.

Quality control and data normalisation was carried out by FIOS Genomics in R/Bioconductor (described in greater detail in Chapter 4.2.3.1). All arrays that passed the quality control including arrays used for cSCC profiling (described in Chapter 4) were normalised together to create a large matrix of normalised expression values.

Differentially expressed probes were identified using the Limma package in Bioconductor. Multiple sampling from the same patient was taken into account in the design matrix for data comparison. P value was adjusted to control for false discovery rate of 0.05. Affyemtrix Array processing that had been applied to this dataset is described in greater details in Chapter 4.2.3.

3.3.2.3. Sun-exposed skin vs. non-sunexposed skin

To validate the hypothesis that the transcriptional profile of non-sunexposed skin differs from sun-exposed skin, I plotted a heatmap of log transformed expression values of all probes present at the microarray (shown in Figure 3.28). Unsupervised hierarchical clustering of the samples shows that while not entirely distinct from each other, the transcription profiles of SE and NSE are sufficiently distinct to support separate comparisons with AK transcription profiles.

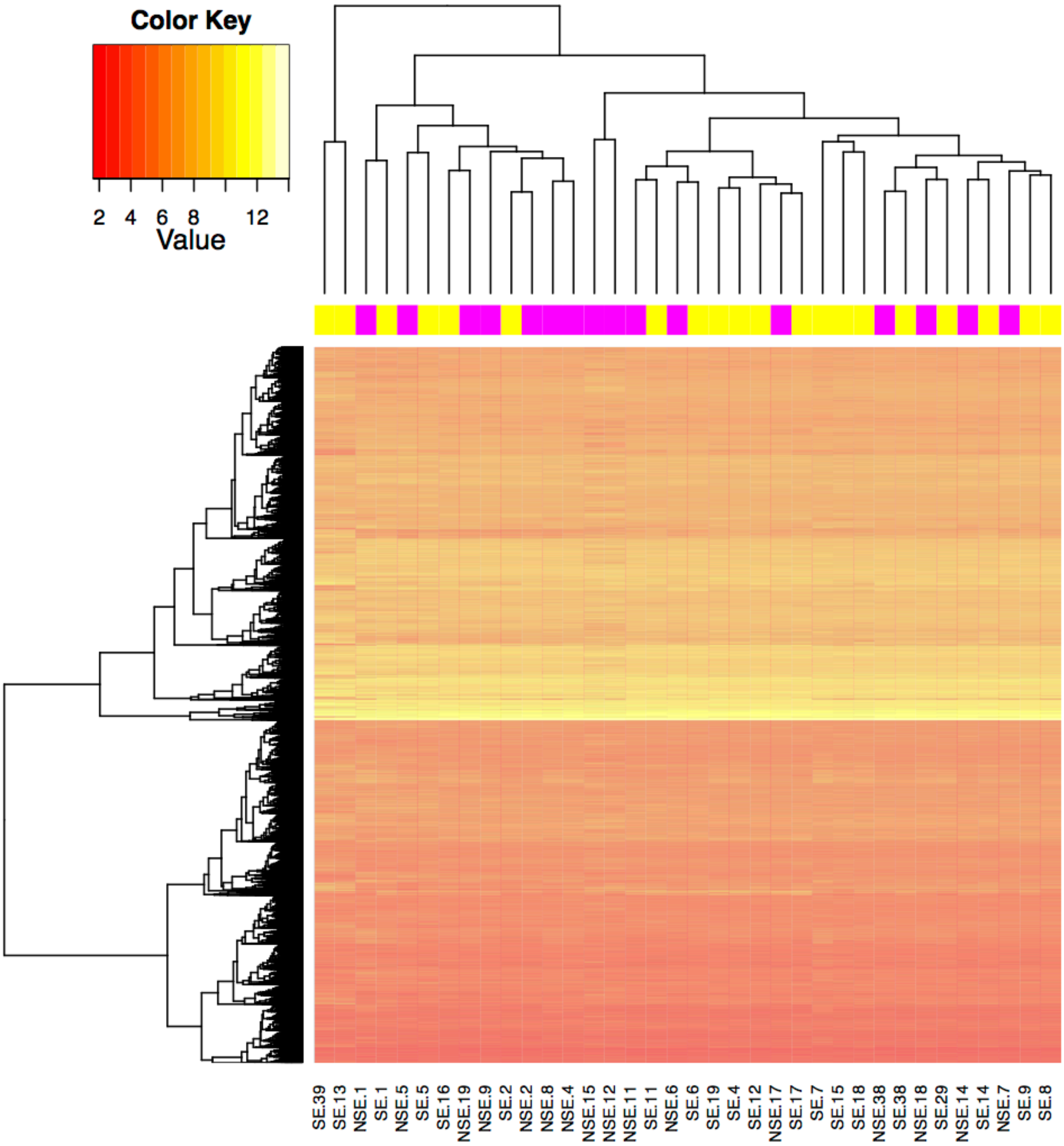


Figure 3.28 Heatmap of unsupervised clustering of NSE and SE samples. Yellow=SE, violet=NSE. Each column represent log2 expression levels of all genes in individual samples

(corresponding numeric values are inferred from the Color Key). In 8 instances, NSE and SE from the same patient clustered together (from the left: numbers 1, 5, 2, 11, 6, 17 38, 14), while in 5 instances skin samples from the same patient did not form a cluster. Additionally, a small cluster of NSE samples is apparent (NSE 2, 8, 4, 15, 12, 11).

The comparison of expression levels between the two tissue types has shown that there were 20 differentially expressed probes at the α level of significance 0.01 (after FDR adjustment for multiple testing). All but three probes were downregulated in NSE compared to SE, and these 20 probes represented 12 unique genes (listed in Table 3.7**Error! Reference source not found.**).

ProbeID	Symbol	Description	Chr	logFC
239260_at	CORIN	corin, serine peptidase	4	-1.531329923
239261_s_at	CORIN	corin, serine peptidase	4	-1.254003291
1563271_at	NA	NA	-	0.931952841
243146_at	NA	NA	-	-1.308232535
203000_at	STMN2	stathmin-like 2	8	-1.610267694
220356_at	CORIN	corin, serine peptidase	4	-1.339283341
236892_s_at	NA	NA	-	-1.201225627
208227_x_at	ADAM22	ADAM metallopeptidase domain 22	7	-0.917965468
230197_s_at	TPPP	tubulin polymerization promoting protein	5	1.157156495
237390_at	NA	NA	-	-1.122921918
212806_at	PRUNE2	prune homolog 2 (Drosophila)	9	-1.55127211
238018_at	FAM150B	family with sequence similarity 150, member B	2	-1.288672304
216887_s_at	LDB3	LIM domain binding 3	10	-0.735158481
244108_at	SYNPO2	synaptopodin 2	4	-1.362182941
203001_s_at	STMN2	stathmin-like 2	8	-1.353975228
241412_at	BTC	betacellulin	4	1.002060116
207092_at	LEP	leptin	7	-1.631393179
231452_at	HRASLS5	HRAS-like suppressor family, member 5	11	-0.79705052
228780_at	NA	NA	-	-0.941476922
228707_at	CLDN23	claudin 23	8	0.892999417

Table 3.7 Differentially expressed genes in NSE compared to SE. Chr=chromosome. LogFC=direction of expression change (negative sign-downregulated in NSE). Adjusted P value for all differentially expressed probes is < 0.01.

I then plotted these probes with corresponding log-transformed expression values in each skin sample. As shown in Figure 3.29, 2 of 3 CORIN probes clustered together, as did both STMN2 probes.

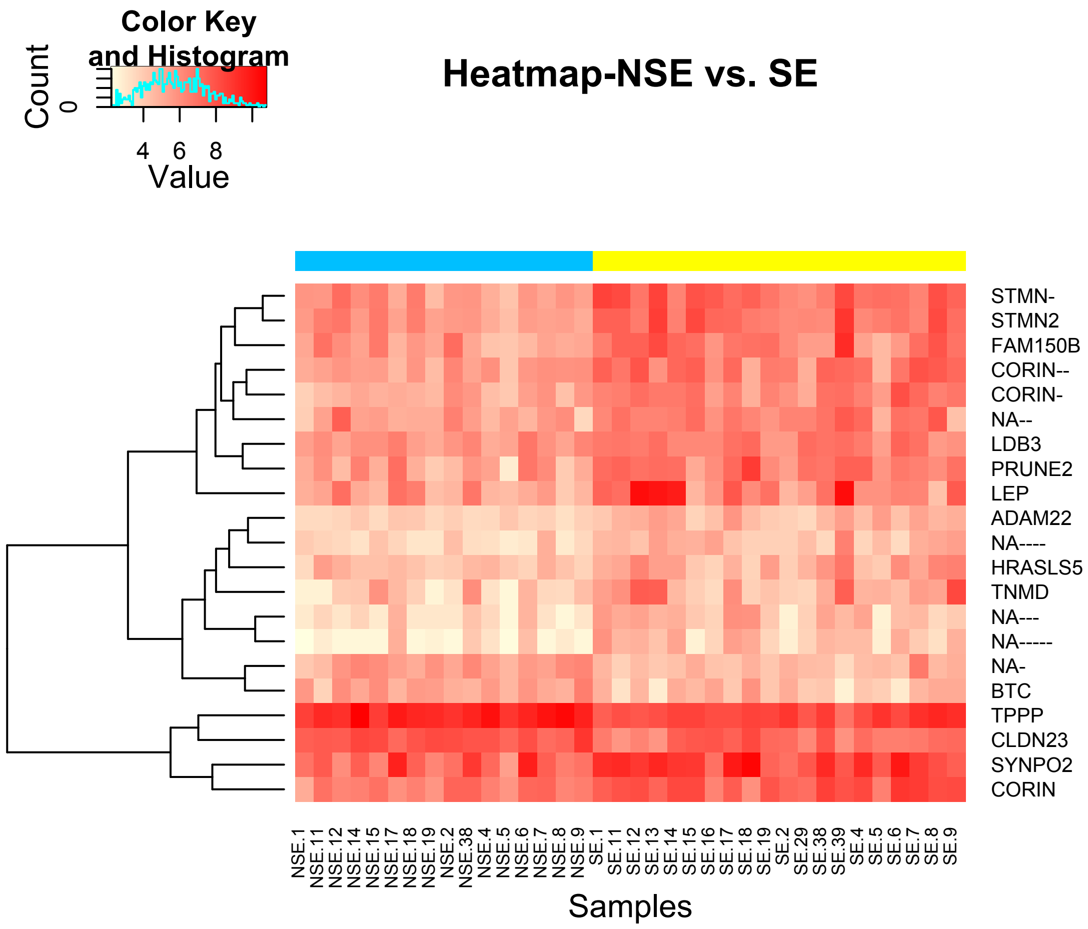


Figure 3.29 Heatmap of log2 expression values of differentially expressed probes with corresponding gene names in NSE and SE. Lighter shades of red indicate down-regulated genes in NSE. Blue=NSE, yellow=SE.

3.3.2.4. Transcription changes between sun-exposed skin and actinic keratosis

A comparison between sun-exposed skin samples and AK revealed a total of 437 differentially expressed probes (p value <0.0001) between the two tissue types. These probes represent 307 unique genes. List of differentially DEG symbols along with chromosomal location, log fold change, p value and description is provided in Appendix 1.

Pathway analysis of this probeset revealed 18 pathways (listed in Table 3.8) that were significantly overrepresented by differentially expressed probes in this dataset (p value <0.05). Many of those point to metabolic disruption in AK with relevance to oncogenic process (PPAR pathway). It is of note that given the complexity of metabolic signalling, many of metabolic pathways are listed because overrepresented probes were functionally involved in multiple metabolic pathways.

Pathway	Sample genes	Adjusted pvalue
3-Chloroacrylic acid degradation	ADH1B, ADH7	<0.0001
1- and 2-Methylnaphthalene degradation	ADH1B, ADH7	<0.0001
Metabolism of xenobiotics by cytochrome P450	ADH1B, GSTM5	<0.0001
Drug metabolism - cytochrome P450	ADH1B, GSTM3	<0.0001
Glycolysis / Gluconeogenesis	ENO1, HK2, ADH1B	<0.0001
Calcium signaling pathway	SLC8A1, PLN, CALML3	<0.0001
PPAR signaling pathway	SORBS1, FABP5	<0.0001
Fatty acid metabolism	ALDH3A2, ADH1B	<0.0001
Tyrosine metabolism	ADH1B, ADH7	<0.0001
Retinol metabolism	ALDH1A2, ADH1B	0.000127
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	DSC2, SLC8A1	0.000159
Adipocytokine signaling pathway	LEP, LEPR,	0.000493
Tetrachloroethene degradation	AKR1B10	0.001094
Olfactory transduction	CALML3, CCLA3	0.002772
RNA polymerase	POLR3G	0.007526
Neuroactive ligand-receptor interaction	NPY1R, LEP	0.012941
Cardiac muscle contraction	TNNC1, SLC8A1	0.027266

Table 3.8 Significantly overrepresented KEGG pathways among differentially expressed probes between SE and AK.

Calcium has been shown to induce apoptosis and found to be dysregulated in prostate cancer based on gene expression profiling (Lee et al., 2003a), I thereby plotted heatmap of calcium signalling pathway probes. As shown in Figure 3.30, most probes were transcriptionally decreased in AK relative to SE, and probes profiling the same gene cluster together with respect to their expression level across all AK and SE samples.

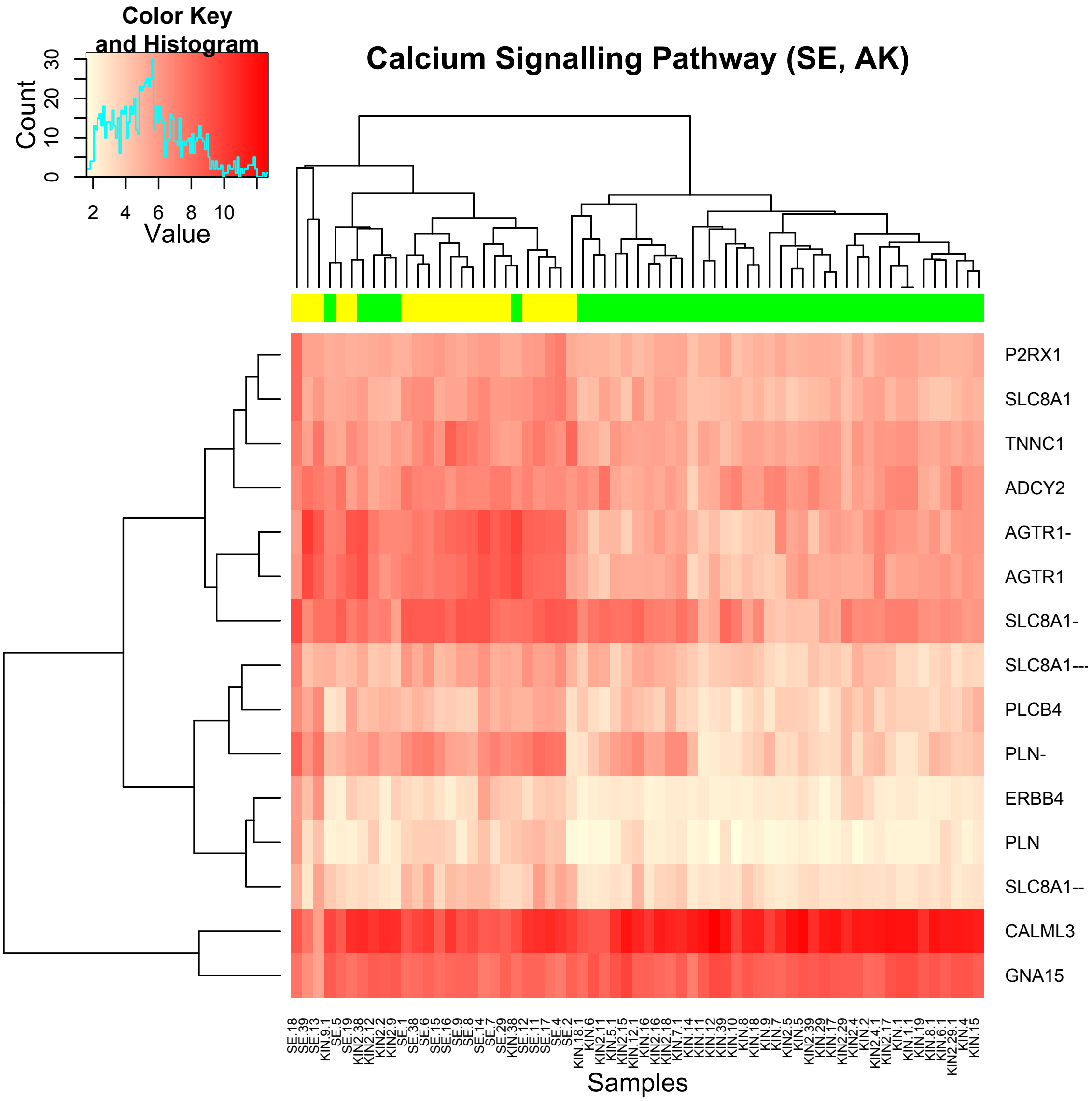


Figure 3.30 Probes involved in Calcium signalling pathway and their log-transformed expression levels in each AK/SE sample. Yellow=SE, green=AK. With the exception of 5 SE samples, the transcriptome corresponding to Calcium signalling pathway molecules segregated the two tissues.

Analysis of GO terms has revealed a total of 166 GO terms to be significantly overrepresented by this probeset (P value <0.01). The full list is shown in Appendix 2, Figure 3.31 shows selected GO terms with a number of corresponding probes.

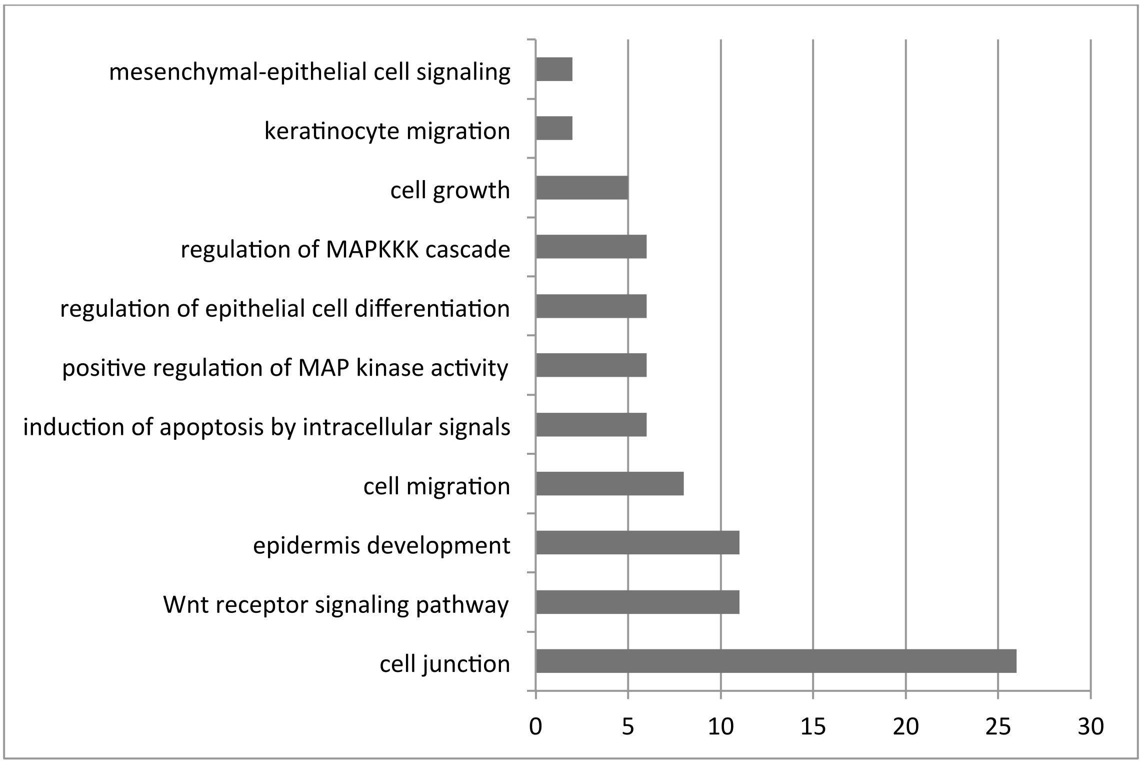


Figure 3.31 Select GO terms and a number of probes annotated with respective term. X axis = number of probes.

3.3.2.5. Transcription changes between nonsun-exposed skin and actinic keratosis

A comparison between nonsun-exposed skin samples and AK has revealed a total of 453 differentially expressed probes (p value <0.0001) between the two tissue types. These probes represent a total of 292 unique genes. List of differentially DEG symbols along with chromosomal location, log fold change, p value and description is provided in Appendix 3.

Pathway analysis of this probeset revealed 18 pathways (listed in Table 3.9) that were significantly overrepresented by differentially expressed probes in this dataset (p value <0.05). Similarly to probes differentially expressed between SE and AK, many are involved in various metabolic pathways.

Pathway	**Sample genes**	**Adjusted pvalue**
3-Chloroacrylic acid degradation	ADH1B, ALDH3A2	9.79E-14
Glycolysis / Gluconeogenesis	ADH1B, ALDH3A2	1.55E-09
Ascorbate and aldarate metabolism	ALDH3A2	5.33E-07
Biosynthesis of unsaturated fatty acids	SCD	3.66E-06
Arginine and proline metabolism	GATM, ALDH3A2	2.47E-05
Fatty acid metabolism	ADH1B, ALDH3A2	4.47E-05
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	CACNB2, SLC8A1	0.000313845

Cardiac muscle contraction	TNNC1, SLC8A1	0.000957663
PPAR signaling pathway	RXRA, SCD, ACADL	0.001083604
Cytokine-cytokine receptor interaction	TNFSF9, CCL27	0.001660786
Streptomycin biosynthesis	HK2	0.001663104
Tetrachloroethene degradation	EPHX2	0.001666642
Limonene and pinene degradation	ALDH3A2	0.00174582
Hypertrophic cardiomyopathy (HCM)	IGF1, SLC8A1	0.002730482
Tight junction	PRKCB, MYH14	0.02073146
1- and 2-Methylnaphthalene degradation	ADH1B	0.02133048
Neuroactive ligand-receptor interaction	LEPR, PARD3	0.02177456
beta-Alanine metabolism	ALDH3A2	0.02416088

Table 3.9 KEGG Pathways dysregulated by differentially expressed probes between NSE and AK.

Probes involved in the Tight junctions pathway are also invoved in processes directly relevant to oncogenesis (such as PRKCB:MAPK pathway). Heatmap depicting log transformed expression levels of probes involved in this pathway shows clear separation of NSE and AK based on this probeset (unsupervised hierarchical clustering) and relatively high differences in absolute expression levels between NSE and AK (Figure 3.32).

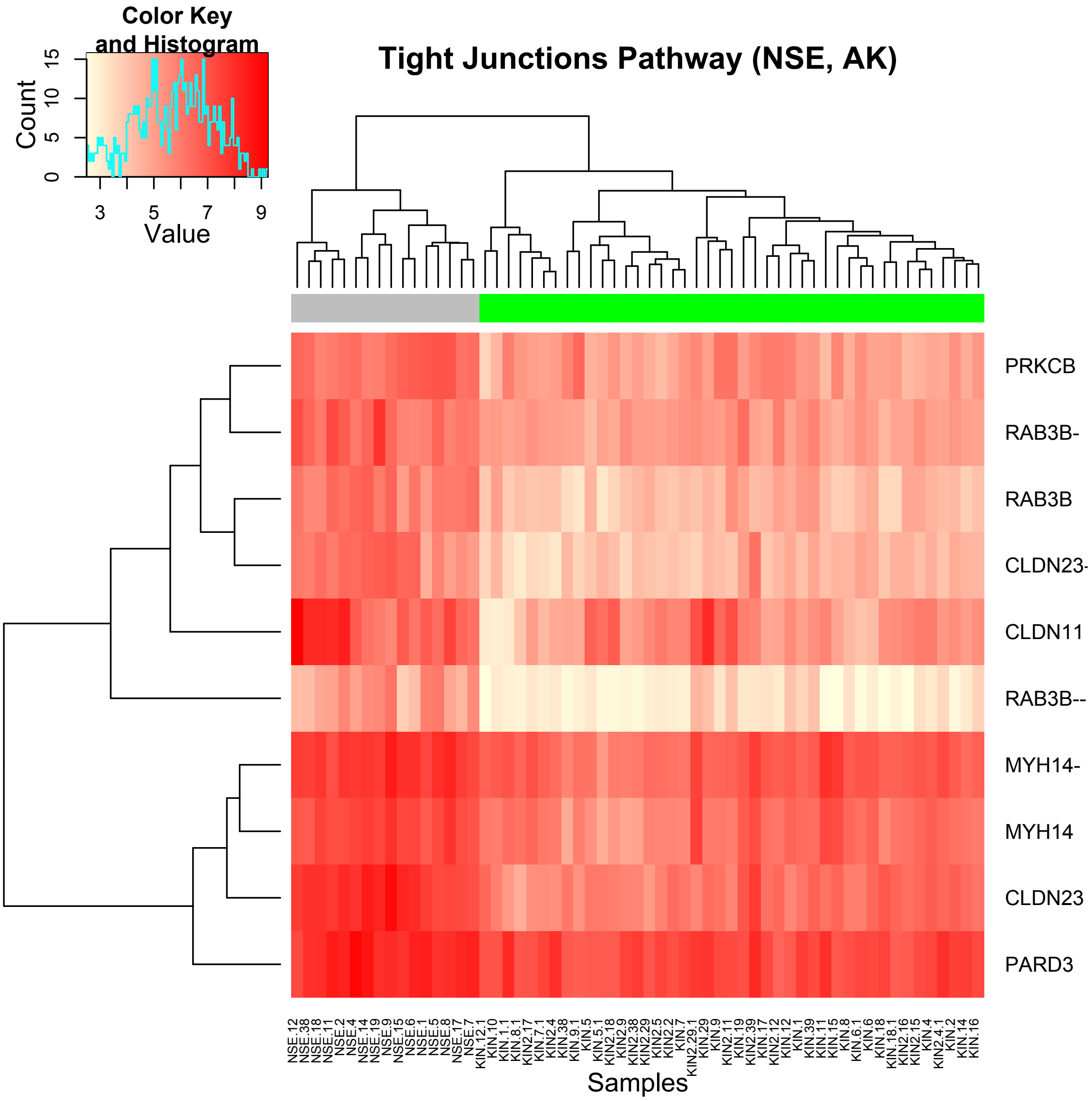


Figure 3.32 Probles invovled in Tight Junctions pathway differentially expressed between NSE and AK. Grey=NSE, green=AK. The transcriptome corresponding to Tight junctions pathway molecules completely segregated the two tissues.

GO term analysis of this set of probes revealed 206 significantly overrepresented GO terms (p value <0.01). The full list if provided in Appendix 4, Figure 3.33 shows a list of selected GO terms with corresponding number of probes.

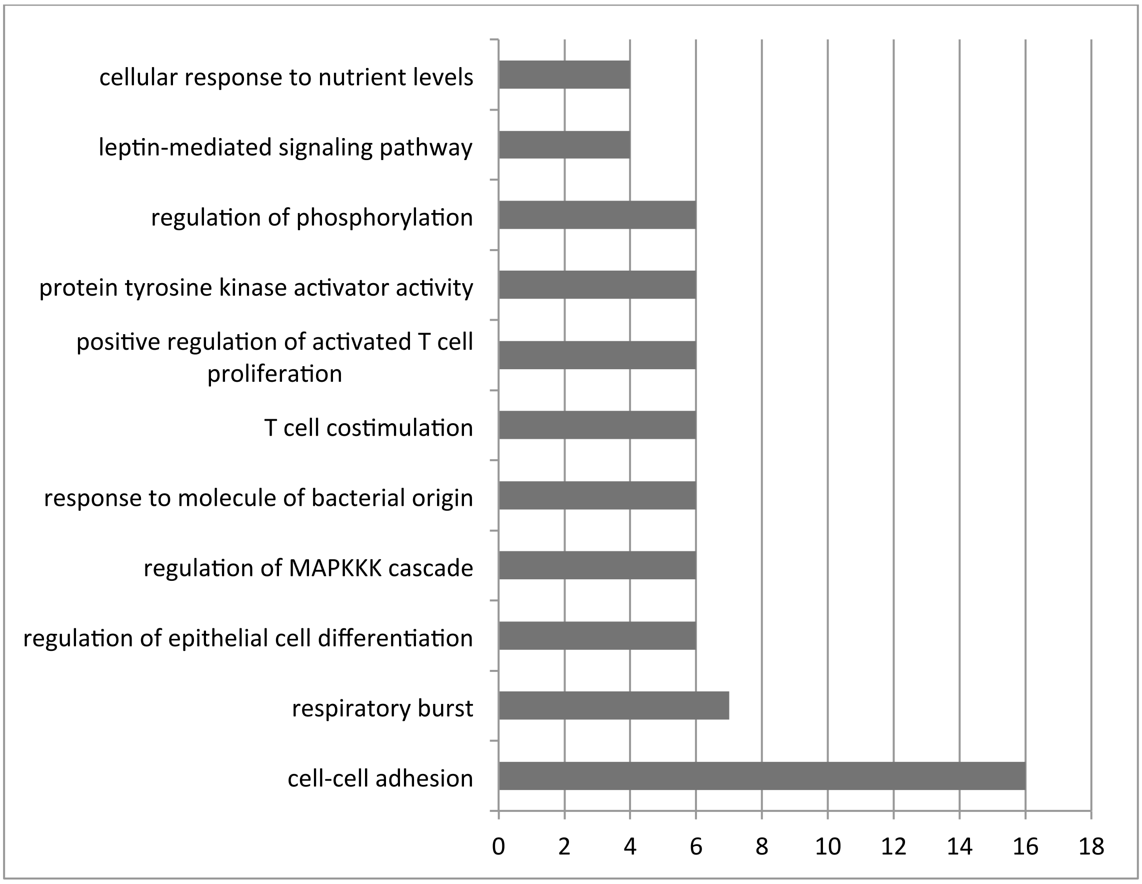


Figure 3.33 Selected GO terms overrepresented among probes differentially expressed between NSE and AK.

3.3.2.6. Overlap between differentially expressed genes, dyregulated pathways and GO terms identified in comparisons between nonsun-exposed skin and actinic keratosis and sun-exposed skin and actinic keratosis

In order to find shared traits of transcriptional differences between skin and AK in general, I compared lists of DEG, dysregulated pathways and GO terms identified in previous comparisons.

A total of 139 unique genes were shared by both comparison groups (45% of SE vs. AK, 47% of NSE vs. AK comparison). Table 3.10 lists 15 genes that are directly involved in mechanisms and processes relevant to oncogenesis or epidermal regulation.

Gene	Name	Comment	Reference
KRT17	keratin 17; keratin 17 pseudogene 3	Upregulated in oral SCC	(Toyoshima et al., 2008)
KRT16	keratin 16; keratin type 16-like	Upregulated in keratoacanthoma	(Ito et al., 2008)
KRT9	keratin 9	Mutated in epidermolytic palmoplantar keratoderma	(Lee et al., 2003a)
KRT6B	keratin 6B		
KRT6A	keratin 6A	Associated with worse prognosis in breast cancer	(Tischkowitz et al., 2007)

S100A2	S100 calcium binding protein A2	Dysregulated in lung SCC	(Strazisar et al., 2009)
S100A9	S100 calcium binding protein A9	Upregulated in cSCC	(Moubayed et al., 2007)
VSNL1	visinin-like 1	TS in lung SCC	(Fu et al., 2008)
S100A8	S100 calcium binding protein A8	Involved in inflammation and cancer	(Gebhardt et al., 2006)
TMC8	transmembrane channel-like 8	Cause of epidermodysplasia verruciformis; linked to HPV and cSCC	(Patel et al., 2008)
TMEM49	transmembrane protein 49	Cell-cell adhesion protein	(Sauermann et al., 2008)
SOX5	SRY (sex determining region Y)-box 5	Overexpressed in glioma	(Ueda et al., 2007)
FOXE1	forkhead box E1 (thyroid transcription factor 2)	Mediates hedgehog signalling in epidermis	(Eichberger et al., 2004)
POU2F3	POU class 2 homeobox 3	Involved in keratinocyte differentiation, enhances HPV replication	(Cabral et al., 2003, Beck et al., 2007, Kukimoto et al., 2008)
DLX2	distal-less homeobox 2	Upregulation is associated with poor prognosis in cancer	(Tang et al., 2013, Yan et al., 2013)

Table 3.10 Differentially expressed genes overlapping between NSE and SE involved in processes relevant to cancer or epidermal regulation. TS=tumour suppressor.

A total of 9 pathways (50%) were shared by both comparison groups, including 1- and 2-Methylnaphthalene degradation, 3-Chloroacrylic acid degradation, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Cardiac muscle contraction, Fatty acid metabolism, Glycolysis / Gluconeogenesis, Neuroactive ligand-receptor interaction, PPAR signaling pathway and Tetrachloroethene degradation. However, many of these pathways were represented by only one or two genes hybridised by abundant number of probes withtin the dataset.

PPAR signalling pathway was represented by entirely different set of genes in both lists: RXRA, SCD and ACADL in NSE vs. AK list, and by SORBS1, PPARD, FABP7 and FABP5 in SE vs. AK list.

GO terms comparison has revealed 72 shared terms (35% of NSE vs AK, 43% of SE vs AK list). These included many immune system related terms (MHC protein binding, T cell costimulation), development (Wnt receptor signaling pathway, epidermis development) and cellular contact (cell junction, cell-cell adhesion). GO terms assigned with most probes are listed in Table 3.11.

GO term	Gene (NSE vs. AK)	Gene (SE vs. AK)
Epidermis development	KRT16, POU2F3, ALDH3A2, KRT9, KRT17, SCEL, EMP1	KRT16, KRT9, KRT17, BNC1, FABP5, POU2F3, ALDH3A2, SPRR1A, SPRR3
Wnt receptor signalling	CD24, WNT16, WNT5A,	WIF1, CD24, WNT5A, DIXDC1,

pathway	GPR177, CTNND1	GPR177
Cell-cell adhesion	CD24, CLEC7A, THY1, BCL2, LMO7, CTNND1	CD24, CLEC7A, BCL2
Cell junction	DSC2, FBLIM1, CLDN23, ENAH, SNTB1, DLG2, GJB2, GJB4, DSG3, CLDN11, GJB6, DENND1A, HOMER1	GJB2, DSC2, ENAH, FBLIM1, SORBS1, GJB6, DSG3, FBXO45, CLDN11, HOMER1, RIMS3, SYNM, DIXDC1, CLCA2

Table 3.11 GO terms with most represented probes in shared by differentially expressed genes in AK compared to NSE and SE.

3.3.2.7. Comparison of transcriptional profile between immunocompromised and immunocompetent patients

Comparisons between OTR and IC patients did not detect any statistically transcriptional differences between the groups in neither normal skin (both sun-exposed and non-sunexposed) or between the AK samples.

3.3.2.8. Validation of transcriptional changes between skin and AK with quantitative PCR

In order to validate genes detected by expression profiling microarray, independent separate panel of 9 AK, 5 SE and 4 NSE samples was used for quantitative PCR (QPCR) analysis (clinical data of this sample set is provided in Appendix 19). Genes included in the validation panel were ID4, CCL24, PTPN21, SLC8A1, CORIN, DLX2, TPPP, NPY1R, ADAM23, HIF1A, COMP, ACVR2A.

Of these 12 genes, 6 were differentially expressed between AK and skin. CORIN was differentially expressed between SE and NSE (not between skin and AK). Figure 3.34 shows a heatmap depicting relative and absolute 2^delta Ct levels of genes that were statistically significantly differentially expressed.

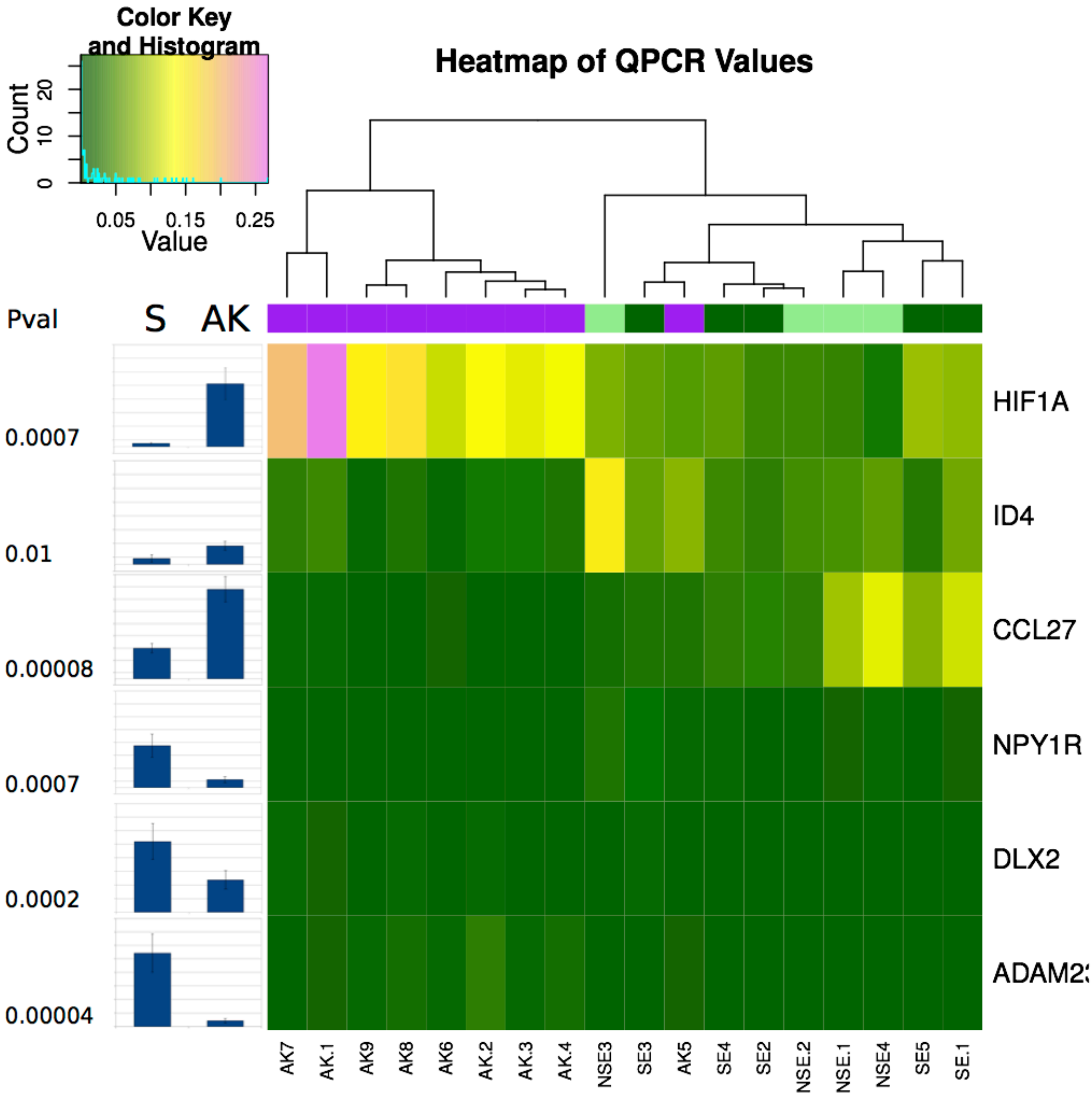


Figure 3.34 Heatmap of normalised Ct values and relative expression of genes used for expressed data validation. Relative expression is depicted next to corresponding gene. Whiskers = standard error, S = skin, AK = actinic keratosis, purple side colour = AK, light green = non-sunexposed skin, dark green = sun-exposed skin. With the exception of 1 AK sample, unsupervised hierarchical clustering of this gene set segregated normal skin from AK.

3.4. Discussion and future directions

3.4.1. Chromosomal aberrations are significantly more frequent in actinic keratosis compared to normal skin

In this chapter, I have shown that AK is a genetically distinct precancerous lesion that accumulates significantly more genomic aberrations compared to normal skin irrespective of skin sun exposure. Additionally, the base-pair size of aberrations in AK is significantly greater compared to aberrations detected in NSE and SE. The majority of SCNA detected in both skin and AK are focal (short), although a full chromosomal arm loss or gain was observed in 6 patients in AK. This is in contrast to a large study of copy-number profiling of various cancers which described focal SCNAs are less frequent compared to whole-arm losses or gains (Gebhardt et al., 2006), and also in contrast to findings in cSCC, in which losses of chromosomal arms were also common findings (Purdie et al., 2009). This may be due to the fact that AK is a precancerous lesion, and that chromosomal aberrations affecting large segments of chromosomes accumulate later in its development. Additionally, this may be to related to AK histological grade.

Focal genomic aberrations detected in multiple patients were found mostly on chromosome 8 and 9; the most common copy-number change in shared SCNAs in AK is monoallelic gain. This finding is largely in accordance with previous studies (Rehman et al., 1994) that detected a high frequency of loss of heterozygosity in AK, and described chromosome 9 as one of the hotspots for LOH. Additional chromosomes frequently affected by SCNA yet containing few if any overlapping regions of genomic imbalance included chromosomes 1, 3, 5, 6 16, and 18. Chromosomal aberrations on these chromosomes with the exception of chromosome 16 were reported in a recent study that used SNP array to detect genomic imbalances in 11 AK samples collected from OTR patients (Patel et al., 2008). This may be a co-incidental observation, or it may indicate that these chromosomes have greater propensity towards genomic instability; possibly due to less resistance to UV-induced damage.

Potential genes affected by imbalances present in AK collected from at least 3 patients include genes located on q-arm of chromosome 9. SYK (spleen tyrosine kinase) and TSC1 (tuberous sclerosis 1) were shown to correspond to amplification hotspots in other neoplasms (Lee et al., 2003a). Additionally, VAV2 (a guanine nucleotide exchange factor) was shown to activate RhoGTPase Rac1, CDC42 and RhoA (Tischkowitz et al., 2007), and the axis of EGFR/VAV and Rac1 activation has been demonstrated as critical for invasiveness in head and neck SCC (Patel et al., 2007). A subset of amplified genes in chromosome 9 SCNAs was described as up-regulated in a series of urothelial carcinoma tumours (Harwood et al., 2000): ASTN2, RXRA, CORO2A, FUBP3 and BRD3, but our data did not find concordant expression levels of these genes. GAS1 and RXRA showed consistency in transcription with gain of genomic material. GAS1 is a Hedgehog pathway member that functions as a tumour suppressor by inhibiting cell proliferation and inducing apoptosis (Nindl et al., 2006). RXRA (retinoid X-receptor α) gene is a member of the PPAR pathway and has been shown to play an important role in epidermial homeostasis and hair growth (Duncavage et al., 2009). Our data warrant further exploration of this gene in AK onset.

3.4.2. Male participants have more genomic aberrations in AK compared to females

Additionally, I detected statistically significant risk difference (42%) in the number of genomic aberration in AK collected from male participants compared to AK collected from female patients. This observation is most likely explicable by greater environmental exposure to UV light among male participants, or by the relatively lower number of female participants in the study. I detected no differences in SCNA size, character or location at gender level.

3.4.3. Transplant patients are significantly more likely to have a loss of genomic material compared to immunocompetent patient

There were no differences in the number of aberrations in skin or AK based on the immune status which is consistent with previous findings (Purdie et al., 2009); however, there are significantly more genomic losses among organ transplant recipients (40%) compared to immunocompetent patients (20%), among whom heterozygous gains predominate. Although one previous study has described LOH to be lower in immunosuppressed individuals compared to immunocompetent patients (Rehman et al., 1997), I did not detect such a trend. This difference is most likely explicable by a greater precision of SNP microarrays in detecting genome-wide regions of genomic aberration. 57% of SCNAs detected across a plethora of human cancer previously corresponded to gains (Ibanez-Ventoso et al., 2008), which almost exactly corresponds with the distribution observed in OTR patients. Thereby the low frequency of gains observed among immunocompetent patients is an unusual finding that requires further validation. Should this observation be validated, it may reflect different immunological properties of cells with loss versus gain of genomic material, or a potential “snowball” effect genomic instability in case of DNA loss that would render such cells more readily as indesirable for an intact immune system. Additionally, given that transplant patients with AK and/or cSCC are on average younger compared to immunocompetent patient, age variation across groups may also be contributing to this difference.

3.4.4. Sun-exposure does not increase the number of genomic aberrations in normal skin

Additionally, I have shown that there are no statistically significant differences between sun-exposed and non-sunexposed skin with respect to the character of genomic aberrations. Although shared regions of genomic changes can be identified in sun-exposed and non-sunexposed skin, such regions of change mostly originate in the same patient, further stressing the over-riding role of individual genetic background.

3.4.5. Important differences in the character of SCNA are evident on individual patient level

The most important factor in the character of SCNAs was patient origin. This finding could be ascribed to individual genomic instability, perhaps due to past history of sun-exposure, and individual propensity to either lose or gain chromosomal segments. Moreover, this may be also explained by chromothripsis of individual chromosomes in individual patients. In corcondace with previous reports, (Zack et al., 2013), individuals with evidence of chromothripsis of one chromosome (such as patient 52, 45 and 15) were also likely to demonstrate chromothripsis of other chromosomes.

Some individual lesions show a very high number of SCNAs and some very little or none (namely in AK), yet clinically these lesions are classified as equivalent - AK. Stratification by histologial grade of AK and patient age, skin type and self-reported history of sun-exposure may clarify reasons for the gargantuan interpatient variation. Yet this observation also suggests that other mechanisms than a chromosomal change in a specific region drive both the onset and potential progression of AK. This may also explain the vast differences in estimates of how many AK actually progress towards cSCC.

3.4.6. Transcriptional profiling of sun-exposed and non-sun exposed skin reveals up-regulation of various cancer-related molecules in sun-exposed skin

Transcriptional differences between sun-exposed skin and non-sunexposed skin detected peptide convertase CORIN to be up-regulated in sun-exposed skin. This enzyme activates atrial natriuretic peptide (ANP) to regulate blood pressure, but also promotes trophoblast invasion and spiral artery remodelling and preeclampsia during pregnancy (Cui et al., 2012). Another up-regulated gene, STMN2, is a target of beta-catenin (Lee et al., 2006a), and shown to play a role in liver carcinogenesis. Addtionally, both betacellulin and TPPP are regulated by the ERK pathway (Hlavanda et al., 2007, Fang et al., 2013), and PRUNE2 is known to play a role in tumour differentiation and prognosis (Fang et al., 2013). Although we did not detect differences at a genetic level between NSE and SE skin, significant transcriptional differences include genes directly relevant in the oncogenic process, suggesting that UV damage leads to transcriptional changes with direct oncogenic potential. The lack of corresponding genetic changes suggests that this may possibly be through epigenetic alterations, or that such transcriptome changes are later reversed through repair mechanisms.

3.4.7. Transcriptional profiling of AK and normal skin reveals differential expression of genes involved in cellular contact, the Wnt and PPAR pathway

A comparison of transcriptional profiling between non-sunexposed skin and AK and sun-exposed skin and AK revealed a total of 292 and 307 genes, with critical oncogenic pathways such as the PPAR pathway essential for epidermal homeostasis (Van Wynsberghe et al., 2011) affected in both genesets.

Additionally, GO term analysis of both gene sets has discovered Wnt receptor signalling pathway dysregulation, in addition to overrepresentation of cell-cell adhesion and cell junction GO terms.

The Wnt receptor signalling pathway has been shown to be dysregulated in colorectal cancer (Budinska et al., 2013), cell-cell adhesion loss is the marker of invasiveness of epithelial cancers, and cell-junction repression is the hallmark of EMT. FBLIM1 is directly involved in cell motility and morphology (Ibragimova et al., 2013), p55 members DLG2 and DLG3 regulate epithelial cell polarity (Caruana, 2002). Interestingly, I also discovered differential expression of GJB2, GJB4 and GJB6, mutations of which are involved in both hearing loss and proliferative skin disorders (López-Bigas et al., 2002). These data suggest that AK is a lesion in which the oncogenic process is initiated by dysregulating cellular adhesion and disrupting conserved regulatory pathways.

3.4.8. Strength and limitations of current study

The strength of our study includes paired genetic analysis of normal skin and AK samples with blood samples, which has been shown to be far superior to even a large group of unmatched controls (Heinrichs et al., 2010), and validation of transcriptional data using qPCR.

Factors that were not directly taken into account in this study include age of participants, skin type, medication and anatomical location of lesions. Lesions were not collected consistently from a single body area, and regional differences in skin may underpin observed changes. Moreover, genomic changes were not validated in a separate set of samples.

Another limitation is the low number of female participants, and the absence of a control group that would help elucidate the functional significance of genomic findings.

Additionally, we could not predict which AK lesions would have progressed to cSCC as all AKs were completely removed for analysis.

In conclusion, these data show that AK is a genetically highly diverse and unstable lesion, which shows much greater proportion of heterozygous losses in organ transplant patients compared to immunocompetent patients. Additionally, the oncogenic process is initiated in the progression from normal skin to AK via disruption of cell contacts and conserved developmental regulatory pathways, such as WNT pathway.

3.4.9. Future directions

SCNAs detected across multiple lesions and the predominance of chromosomal gains in immunocompetent individuals need to be validated in an independent set of AK sampels collected from both immunocompromised and immunocompetent patients using a different technique, such as FISH. A larger study with more female participants, control lesions from all participants with AK and control group will provide insight into the significance of focal SCNA in normal skin of AK patients, and additional non-precancerous skin lesions such as warts and seborrhoeic keratosis representing hyperproliferative but benign control lesions. Such study would provide distinction between SCNA that are entirely benign, relevant for keratinocyte proliferation but not conferring malignant potential, and AK-specific.

4. Gene Transcription Profiling of Actinic Keratosis and Cutaneous Squamous Cell Carcinoma

4.1. Introduction

As discussed in detail in Chapter 1, previous investigations focusing on unweaving the genetic background of cSCC have described various karyotypic changes in this tumour, yet with little consensus. Recent investigations from our laboratory using single nucleotide polymorphism microarrays revealed loss of heterozygosity at 3p and 9p sites as the most common genetic events in 60 cSCCs. In addition, this study has shown that clinically relevant histopathololological subtypes of cSCC (well differentiated, moderately differentiated, poorly differentiated) exhibit distinct SCNA changes (Purdie et al., 2009). However, none of the previous studies has provided conclusive results with respect to cSCC tumourigenesis nor pointed firmly towards novel therapeutic options.

Several previous studies have used microarrays to characterise the cSCC transcriptome, but the results of individual studies are largely conflicting (Van Haren et al., 2009). Reasons for this may include the use of different methodologies or statistical approaches, as well as the heterogenous nature of cSCCs, which commonly show leukocyte and stromal infiltration, making molecular studies on tumour tissue more challenging and general findings derived from crude tumour tissue less interpretable.

The purpose of this investigation was to compare the transcriptome profile of tumour cells isolated from cSCC tumours and dysplastic keratinocytes isolated from AKs, and to identify differentially expressed genes (DEG), potentially dysregulated cancer pathways, and enriched Gene Ontology terms within the differentially expressed genes set. In addition, I aimed to examine the association between expression profiles and tumour histopathology, patient gender and immune status, and concordance of expression changes with known genetic aberrations. I also hypothesised that integration of these data with SNP array profiling might identify potential targes for epigenetic regulation of cSCC carcinogenesis. Finally, I hypothesised that cSCC acquire stem cell-like transcriptional profile due to a degree of dedifferentiation present in those tumours, and that this difference would be more pronounced in less differentiated subtypes (MD and PD tumours).

4.2. Materials and Methods

4.2.1. Histopathology of samples

Haematoxylin-eosin (H&E) stained slides of tumours included in the study were evaluated by a histopathologist[12] to ensure intra-observer consistency. Based on histological characteristics, tumours were designated as well-differentiated (WD), well-to-moderately differentiated (WMD), moderately-differentiated (MD), moderately-to-poorly differentiated (MD/PD) and poorly differentiated (PD). Additionally, two cSCC nodal metastases were included in the expression data series.

4.2.2. Expression microarray hybridisation

Ten nanograms of RNA were used as a template for whole transcriptome amplification and cDNA systhesis using the NuGen WT-ovation Pico RNA Amplification System according to the manufacturer's instructions (NuGen, USA). Five micrograms of amplified cDNA were labelled with the FL-ovation Biotin Kit (NuGen, USA) and hybridised to the Affymetrix HGU133 Plus 2.0 microarrays (Affymetrix, USA), comprising 54,675 probes.

4.2.3. Statistical analysis

4.2.3.1. Quality control and data normalisation

Quality control and data normalisation was conducted by FIOS Genomics. Standard quality control checks were performed on all .CEL files using the arrayQualityMetrics package in Bioconductor. Arrays were scored on the basis of 6 metrics: maplot, boxplot, heatmap, spatial, rle, nuse. Arrays not passing two or more of the quality control checks were removed; 5 of the 40 arrays (2 AKs, 3 cSCCs) failed the quality control by these criteria and were removed from further downstream analysis. In addition, any features present on fewer than 20% of the arrays were removed, leaving 36,055 features, which were then normalised using robust multi-array averaging (RMA). Array features with an inter-quartile range (IQR) of <0.5 log2 intensity units were removed, leaving a total of 24,573 features for downstream analysis. All expression arrays (skin, AK, cSCC, nodal metastases) were normalised together and subsets of the normalised matrix were used for subsequent tissue-specific analyses.

4.2.3.2. Detection of Differentially Expressed Genes

Detection of differentially expressed genes was conducted by FIOS Genomics. Empirical Bayesian statistics were used to identify differentially expressed genes using the Bioconductor Limma package, taking into account multiple sampling from the same patient. Vertical p value

---

[12] Professor Rino Cerio, Department of Dermatology, Barts and the London NHS Trust, London, UK

adjustment for multiple testing was applied to control for false discovery rate of 0.05. Genes with adjusted p value < 0.0001 were considered to be differentially expressed genes (DEG).

4.2.3.3. Integration of expression and SNP microarray data

Previously published SNP microarray dataset (Purdie et al., 2009) was analyzed with respect to loss of heterozygosity or gain of allelic copy number in defined chromosomal locations. Twenty-one samples that were profiled in the current study were also included in the previous SNP experiments. SCNA changes that occurred in at least 10% of the samples (6/60) were considered a consistent region of allelic imbalance (Walker et al., 2006). The chromosomal position of each DEG was compared with the coordinates of consistent regions of allelic imbalance, and those genes located within this region that showed a fold-change consistent with the imbalance (decrease in expression in case of deletion, increase in expression in case of gain), were considered to correlate with SCNA changes.

4.2.3.4. Gene set Enrichment

Gene set enrichment analysis searches for common biological functions or pathway dysregulation for a given geneset. I performed pathway analysis in order to identify pathway(s) predominantly disrupted in the progression from AK to cSCC. Each probe was annotated with respective pathways it interacts with according to the Kyoto Encyclopedia of Gene and Genomes (KEGG) database. A chi-square test was performed on DEG using all probes as the background. Similar approach was applied to Gene Ontology enrichment analysis. Probes were annotated with respective GO terms by interrogating the GO database (www.geneontology.org) and chi-square test was applied analogously.

4.2.3.5. Cluster analysis of expression profiles

Hierarchical clustering of the samples was performed to detect profile correlation between samples on the DEG using distance matrix computation in R, visualising the results with heatmap and heatmap.2 functions in the R software. Clustering based on genome-wide transcriptional profiling was conducted using Pearson correlation and the VSN package.

4.2.3.6. Correlation of expression data clinicopathological features by multiple regression analysis

For statistical purposes, MD/PD and PD tumours were grouped together. The top 200 differentially expressed probes from each tumour were combined according to histological stratification and ANOVA testing was applied using Graphpad software. Furthermore, multiple regression analysis assessing the expression profile and its correlation with histological subtype, immunity status and gender was applied using R. Correlation of histological typing with expression profiles was further validated using pvclust package (**http://www.is.titech.ac.jp/~shimo/prog/pvclust/**) in R.

4.2.3.7. Quantitative real-time PCR

QPCR was conducted in collaboration with Dr Karin Purdie: altogether, 13 genes were assessed using quantitative real-time PCR. Quantitative RT-PCR was performed on a separate panel of samples to those used for expression microarray analysis. Total RNA was extracted directly from 32 fresh frozen biopsies from 5 normal skin samples, 9 AK, 9 WD SCC and 9 MD SCC using the RNeasy Mini kit with on-column DNAse digestion (Qiagen, United Kingdom) according to the manufacturer's instructions. First strand cDNA was synthesised from 500ng RNA using Superscript III First-Strand Synthesis SuperMix, with a combination of 2.5 uM oligo dT and random primers (2.5ng/ul) (Invitrogen, UK). The cDNA was diluted 1:10 prior to real-time quantitative PCR (QPCR) analysis. QPCR reactions were run in triplicate on the ABI 7500 Real time PCR machine using 1ul of diluted cDNA and Brilliant II SYBR Green QPCR Mastermix (Stratagene, UK) with the two-step cycling protocol recommended by the manufacturer. The PCR primers were intron-spanning with the exception of those for *JUN*, which has a single exon. Relative expression data were calculated by the equation $2^{\Delta\Delta Ct}$, using the endogenous control gene RPS9 as the internal reference. Data was normalised to the mean Ct value from 5 normal (non-lesional) skin samples. For each sample, cDNA synthesis was also performed in the absence of reverse transcriptase to exclude the possibility of genomic contamination.

QPCR data was normalised using normal skin as a reference. Subsequent values were tested for normality using Shapiro-Wilk test in R and difference between AK and SCC was evaluated using Mann-Whitney U test. The α-level was set at $P<0.05$.

4.2.3.8. Comparison with previously published dataset

Our raw expression data were compared with previously published mesenchymal stem cells expression data (Tsai et al., 2007) and a comprehensive dataset assessing expression levels of various cutaneous samples (Riker et al., 2008).

Additionally, a published set of 15 skin samples (11 skin samples (5 pooled), 5 actinic keratosis (2 pooled), and 5 invasive cSCC (Nindl et al., 2006) was used for further validation of the results.

4.2.4. Experimental design and comparison strategy

In total, 30 fresh frozen cSCCs and 10 fresh frozen AKs were subjected to microarray expression profiling. These samples were obtained from a total 32 patients. In 9 cases, cSCC samples were matched with AK from the same individuals.

I then compared the genome-wide expression profile of 27 cSCC (13 WD, 1 WD/MD, 7MD, 4 MD/PD, 2 PD) and 8 matched AKs that passed the stringent quality control (Table 4.1). The samples were derived from 12 immunocompetent patients, 22 organ transplant recipients (OTR) and 1 patient with chronic lymphocytic leukaemia, reflecting the diverse spectrum of patients affected with cSCC.

4.3. Results

4.3.1. Genome-wide transcription profile of AK and cSCC

Genome-wide transcription profile of a total of 40 fresh-frozen, laser-capture microdissected samples was obtained using HGU133 Plus 2.0 expression microarray. Basic characteristics of the samples are shown in Table 4.1.

Sample	Histology	Immune Status	Differentiation	QC
1	AK	OTR	/	Pass
2	AK	OTR	/	**Fail**
3	AK	OTR	/	Pass
4	AK	IC	/	Pass
5	AK	IC	/	Pass
6	AK	OTR	/	Pass
7	AK	OTR	/	Pass
8	AK	OTR	/	Pass
9	AK	IC	/	Pass
10	AK	OTR	/	**Fail**
11	SCC	OTR	WD	**Fail**
12	SCC	OTR	WD	Pass
13	SCC	OTR	MD	**Fail**
14	SCC	OTR	WD	Pass
15	SCC	OTR	WD	Pass
16	SCC	OTR	WD	Pass
17	SCC	OTR	WD	Pass
18	SCC	IC	WD	Pass
19	SCC	OTR	WD	Pass
20	SCC	IC	WD	Pass
21	SCC	IC	MPD	Pass
22	SCC	OTR	MPD	Pass
23	SCC	IC	WD	Pass
24	SCC	IC	WD	Pass
25	SCC	OTR	WD	Pass
26	SCC	OTR	MPD	Pass
27	SCC	OTR	PD	Pass
28	SCC	OTR	PD	Pass
29	SCC	OTR	WMD	Pass
30	SCC	CLL	MD	Pass
31	SCC	OTR	WD	Pass
32	SCC	OTR	MD	Pass
33	SCC	OTR	MD	Pass
34	SCC	OTR	MD	Pass
35	SCC	IC	MD	Pass
36	SCC	OTR	MPD	Pass
37	SCC	IC	WD	Pass
38	SCC	OTR	WD	**Fail**
39	SCC	IC	MD	Pass
40	SCC	IC	MD	Pass

Table 4.1. Basic characteristics and quality control outcome of AK and cSCC clinical samples hybridised to gene expression microarray. 9 AK are matched with cSCC, samples originated in a total of 32 patients. QC=quality control, OTR=organ transplant recipient, IC=immunocompetent.

WD=well-differentiated, WMD=well to moderately differentiated, MD=moderately differentiated, MPD=moderately to poorly differentiated, PD=poorly differentiated cSCC.

In order to determine the level of correlation between the transcriptional profile of two sample groups (AK and cSCC), I carried out unsupervised hierarchical clustering analysis of all normalised probes present on the array that were expressed above the background (passed the flitering step). Figure 4.1 demonstrates that genome-wide transcription profile of AK and cSCC is generally sufficiently different to warrant comparison on gene level in order to detect differentially expressed genes. 1 AK sample derived from a cardiac transplant patient showed rather distinct transcription profile and clustered aside both groups. 1 cSCC samples clustered among 4 AK samples (well-differentiated sample derived from an immunocompetent patient), and 1 AK formed a cluster with 3 cSCC.

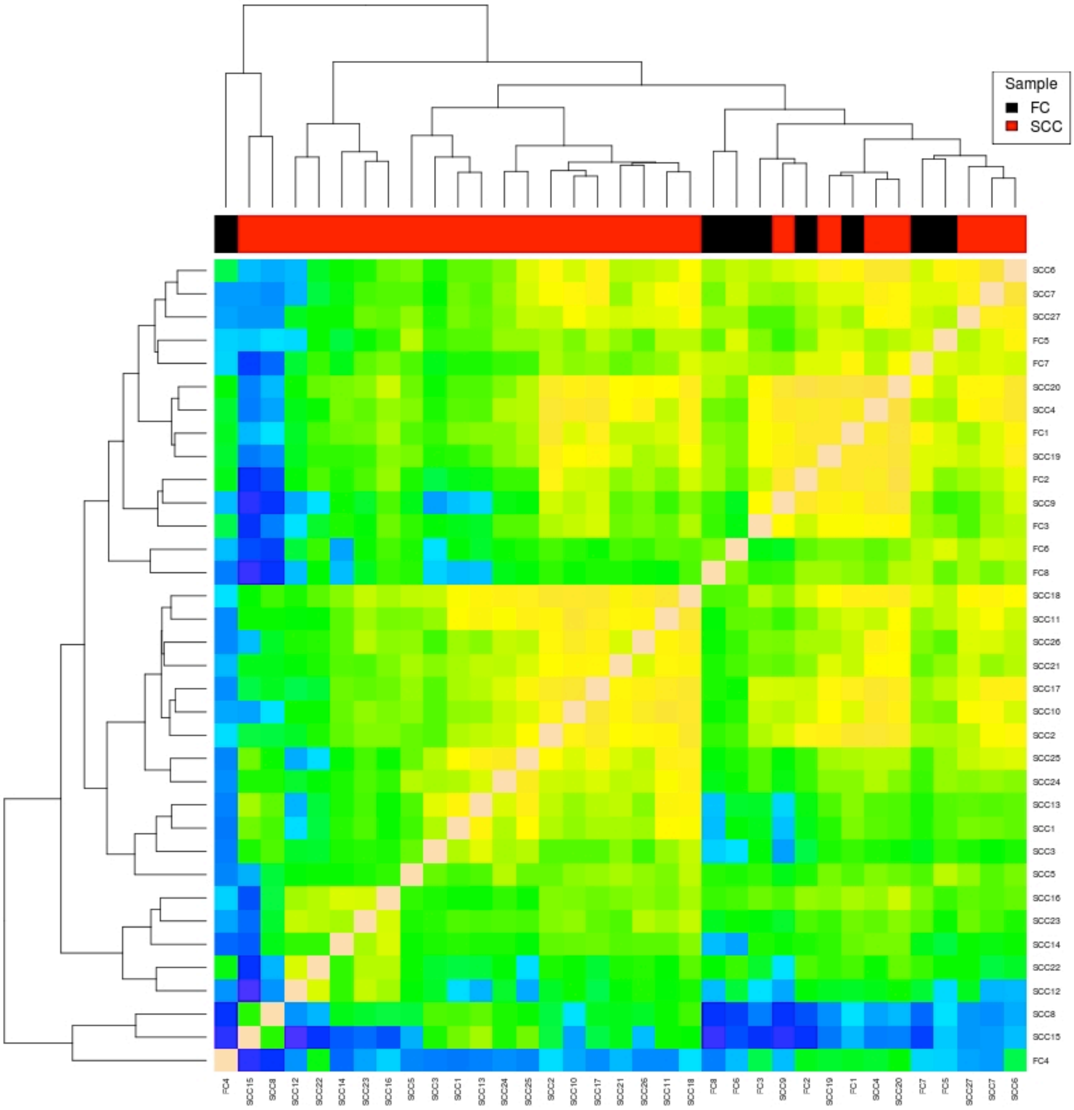


Figure 4.1 Hierarchical clustering of AK and cSCC transcription profiles based on Pearson's correlation of all features present in the quantile-normalised matrix. This approach of assessing

pan-transcriptomic similarity between samples in its essence reveals the first component of principal component analysis. Majority of tumours cluster together, with the exception of one well-differentiated tumour forming a cluster with 4 AK samples. One AK sample has a transcriptional profile distinct from the remaining samples. FC=AK.

4.3.1.1. Differentially Expressed Genes between cSCC and AK

Analysis of differential gene expression revealed 508 probes to be differentially expressed with p value < 0.0001. Unsupervised clustering of those probes showed almost complete separation of AK and cSCC (Figure 4.2). This probe-set represents 346 differentially expressed genes (excluding genes lacking annotation). Of these genes, 147 (42%) were down-regulated in cSCC compared to AK, while 199 genes (58%) were up-regulated.

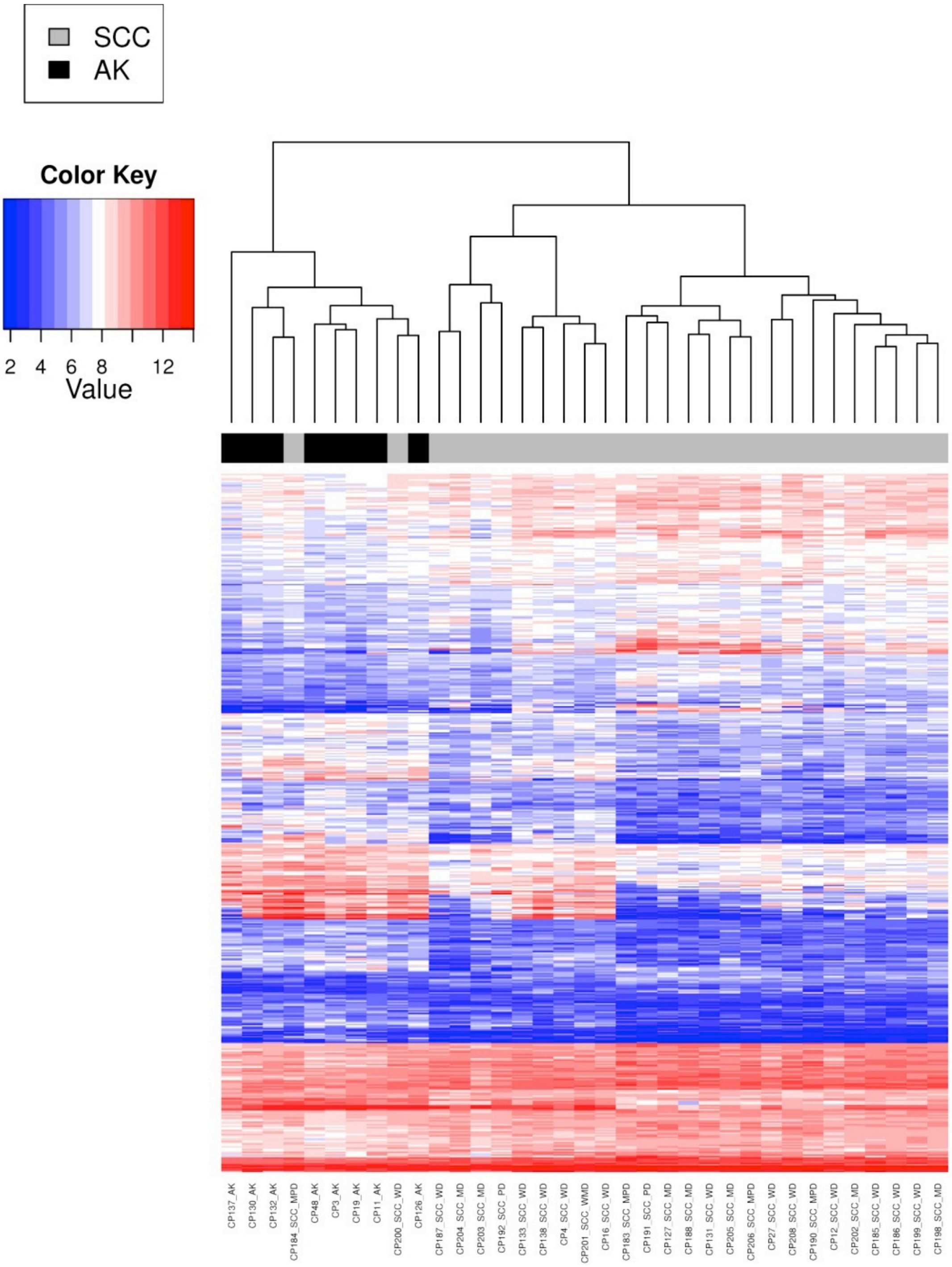


Figure 4.2 Unsupervised hierarchical clustering of log2-transformed expression values of top 508 dysregulated probes in AK and cSCC samples. With the exception of 2 tumours (1 MPD and 1 WD cSCC), this probeset leads to a clear segregation of AK and cSCC.

4.3.1.2. Upregulated genes are involved in cancer invasion, metastasis, epithelial-mesenchymal stem cell remodelling and inflammation

Heatmap of top 50 upregulated genes present among 346 differentially expressed genes based on log fold-change is shown in Figure 4.3.

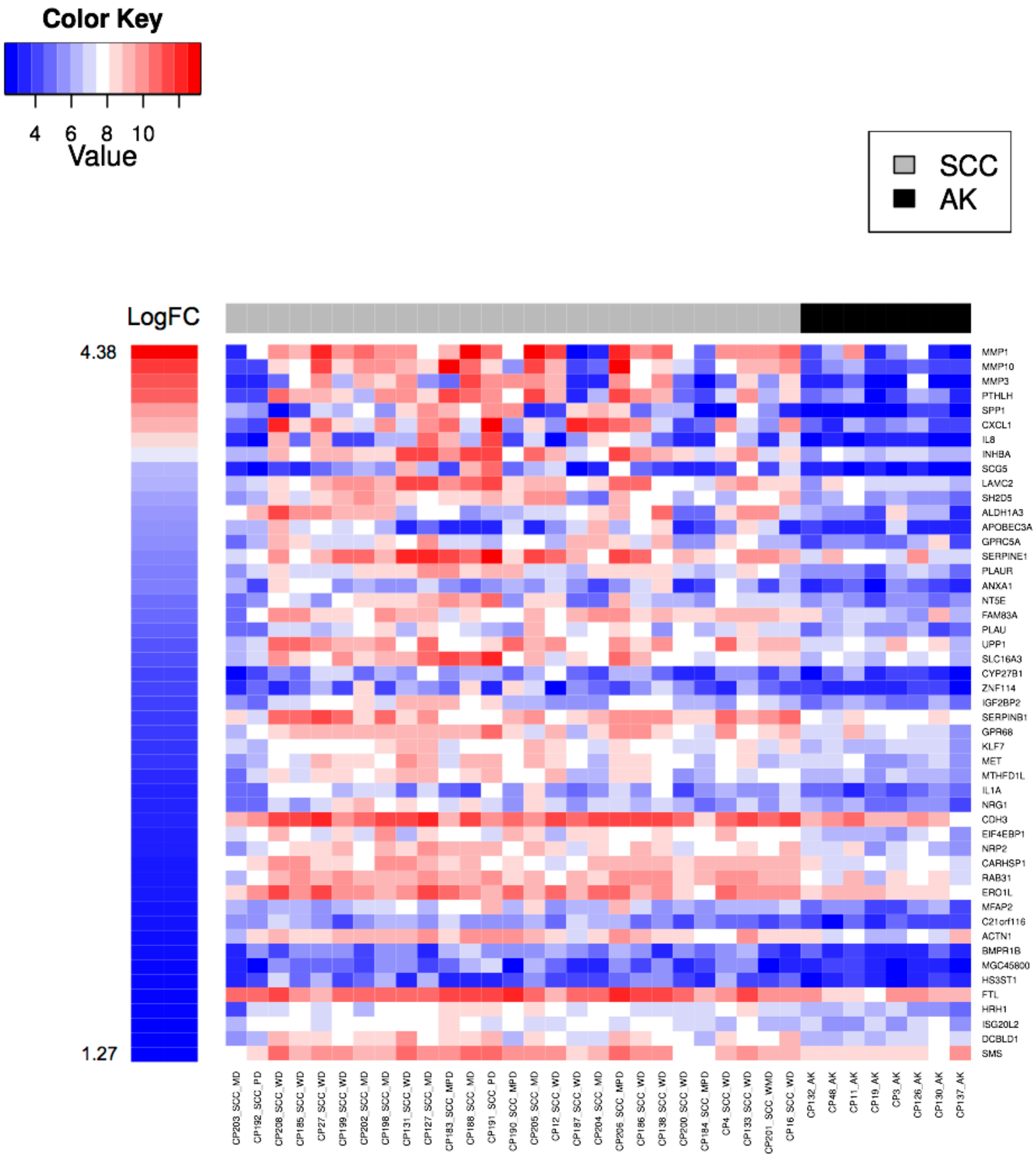


Figure 4.3 Heatmap of top 50 upregulated genes based on log fold-change. While differences in expression level across individual samples are discernible, namely in the two left-most MD and PD cSCC samples, the expression levels of the genes show a large consistency across the two tissue types.

The group of top up-regulated genes contained several proteinases and other factors known to be involved in cancer invasion and metastasis, including: metalloproteinases (MMP1, MMP10, MMP3), which have been demostrated to play a critical role in cancer initiation and metastases (Suomela et al., 2009); parathyroid hormone-like hormone (PTHLH) recently shown to be a candidate for breast cancer susceptibility (Ghoussaini et al., 2012); osteopontin (SPP1) whose over-expression has been demonstrated to correlate with invasion in ovarian cancer (Matsuura et al., 2010) and breast cancer (Cook et al., 2006); and LAMC2 shown to enhance invasion in various cancers (Tsubota et al., 2010, Yamamoto et al., 2009). Other up-regulated genes included INHBLA, a poor-prognosis marker in gastric cancer (Wang et al., 2010), and MET, a putative oncogene.

The up-regulation of SERPINE1, PLAU and PLAUR, which are involved in angiogenesis, response to hypoxia, keratinocyte migration (Villar et al., 2010) and extracellular matrix remodelling, was of significance, since SERPINE1 binds to PLAU, thus activating PLAUR. This complex subsequently interacts with various integrins (Tsai et al., 2007), whose down-stream effectors activate ACTN1, also highly-upregulated in our dataset (Figure 4.4). This mechanism is involved in matrix remodelling in mesenchymal stem cells. Interestingly, significantly up-regulated SERPINE1, ANXA1, INHBA were shown to be highly overexpressed in mesenchymal stem cells, while another highly up-regulated molecule, HRH1 is expressed in differentiating mesenchymal stem cells (Pochampally et al., 2007), and IGF2BP2, a stem cell marker (Fong et al., 2011).

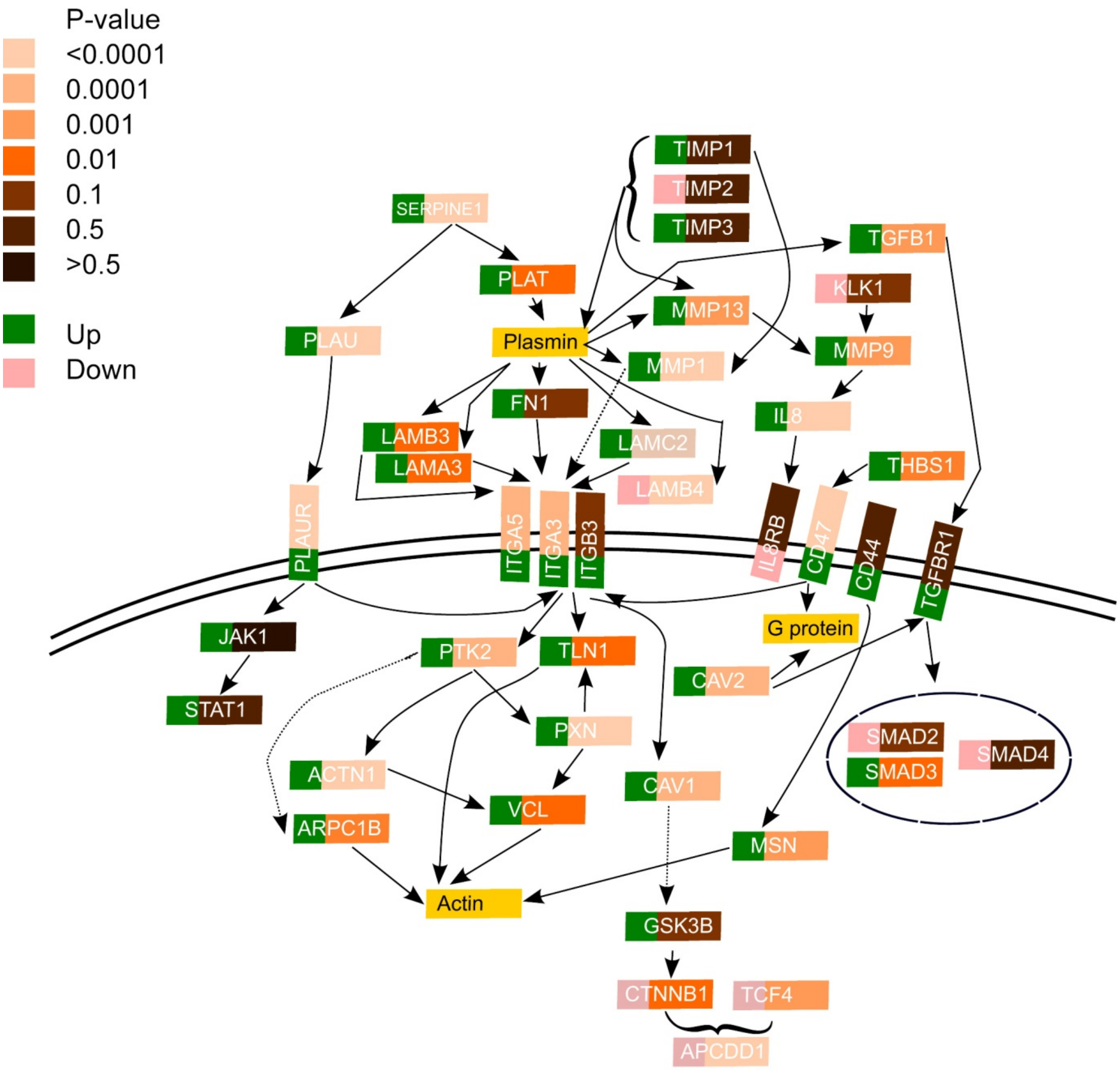


Figure 4.4 Summary of ECM regulation and expression levels of relevant molecules in our dataset.

In addition to molecules described above, several pro-inflammatory cytokines were significantly over-expressed in our dataset, including CXCL1, an oncogene linked to several cancers including melanoma (Dhawan and Richmond, 2002), and IL8, whose overexpression has been linked to more aggressive profile of prostate cancer (Manna et al., 2013). Many genes involved in stromal progression of cancer and stromal remodelling were also upregulated as expected.

The average expression of MMP1, MMP10, MMP3, INHBA, PTHLH and LAMC2 on the array is shown in Figure 4.5, and indicates that there is a gradient of expression across histological cSCC subtypes in LAMC2 expression levels, and that MPD/PD tumours show the highest range of expression levels of these molecules (Figure 4.5), indicating greatest transcriptional hererogeneity within these tumours.

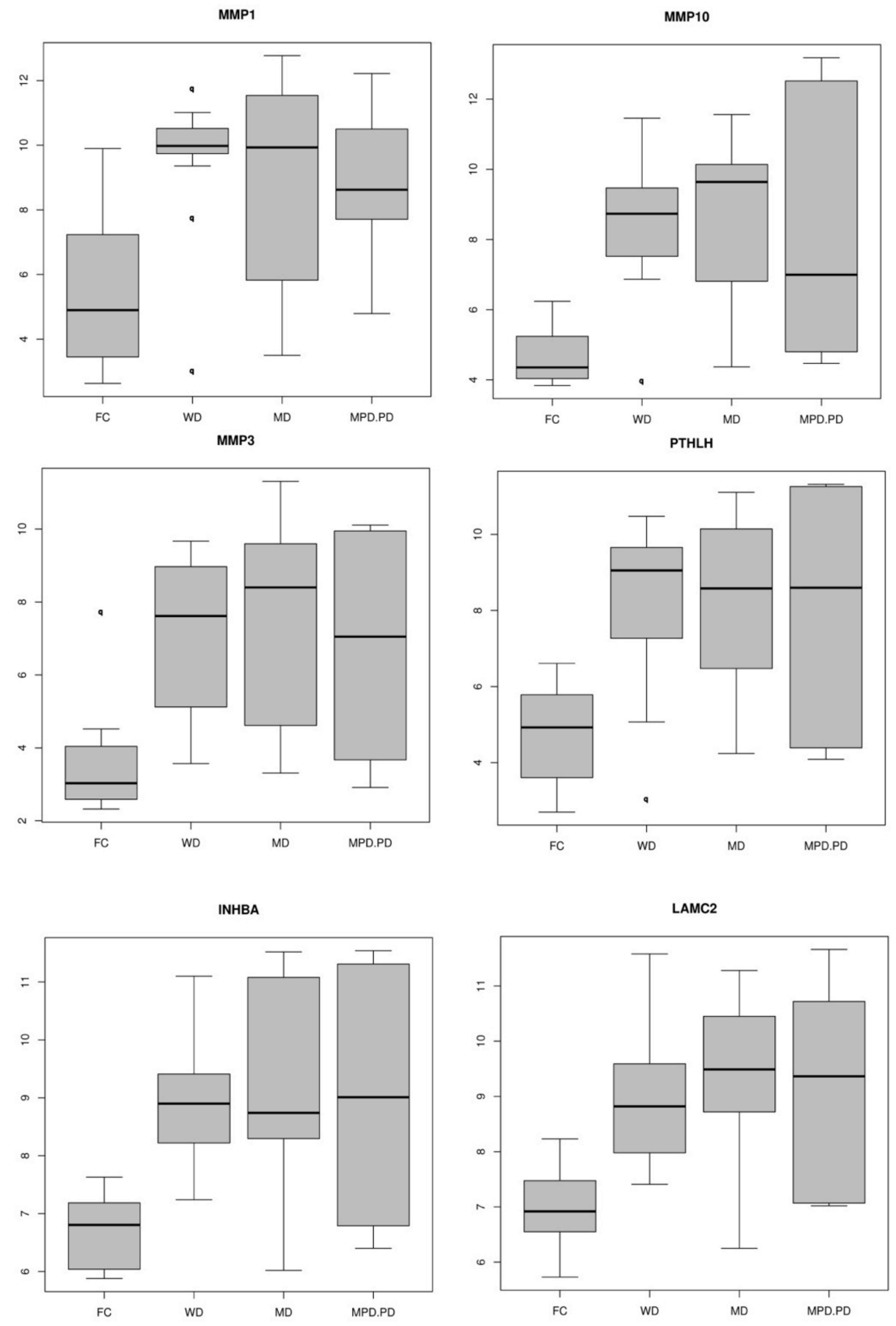


Figure 4.5 Expression of selected genes across histological subtypes of cSCC. A gradient of expression levels is apparent across higher histological specimen for LAMC2. FC=AK.

4.3.1.3. Downregulated genes include genes associated with epidermal differentiation, tumour suppressors, and epidermal stem cells.

Heatmap of top 50 downregulated genes based on log fold-change is shown in Figure 4.6. It shows that a subset of 7 cSCC showed level of expression of these genes similar to AK (with one exception, all were WD tumours).

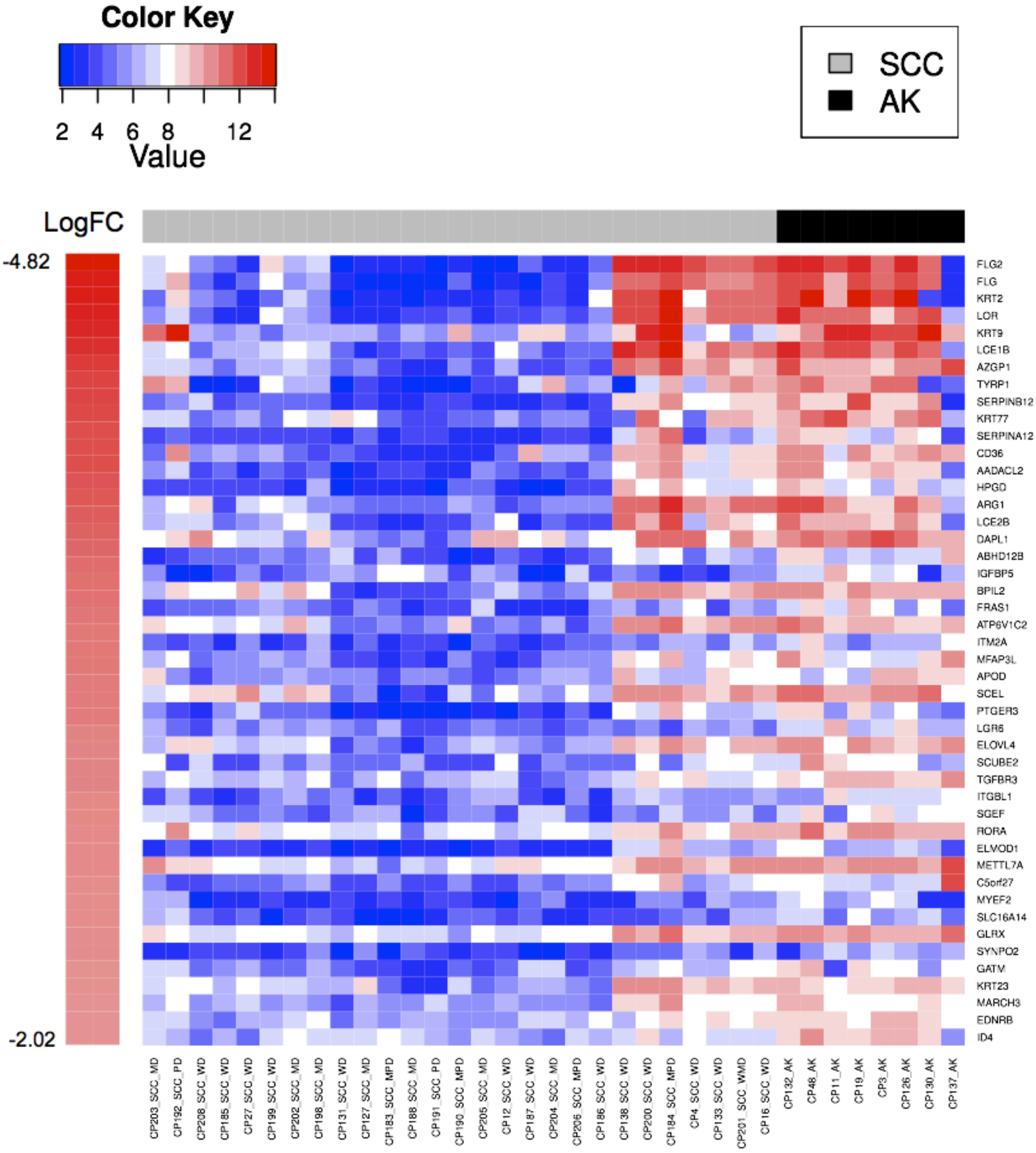


Figure 4.6 Top 50 down-regulated genes based on log fold-change. 7 cSCC tumours (5 WD, 1 WMD and 1 MPD cSCC) adjacent to the AKs demonstrate expression levels of these genes similar to AKs.

Down-regulated genes included several markers and inducers of epidermal differentiation, such as KRT9, FLG, FLG2, LOR, LCE1B, LCE2B, SERPINA12 and SCEL. We also found down-regulation of several tumour suppressor genes including AZGP1, shown to act as a tumour suppressor in prostate cancer by inhibiting TGB-beta mediated ERK signalling (Kong et al., 2010); SCUBE2, also involved in hedgehog signalling (Hollway et al., 2006); SERPINB12 which is down-regulated in oral squamous cell carcinoma (Shiiba et al., 2010) and highly expressed in uppermost layer of normal epidermis; HPGD, a known tumour suppressor in various cancers (Mohamed et al., 2011); IGFBP5, a tumour suppressor in head and neck SCC (Hung et al., 2008); FRAS1 involved in epidermal adhesion (Short et al., 2007). LGR6, a marker of epidermal stem cells (Snippert et al., 2010), was also down-regulated in our study. The average expression of LOR, HPGD, FLG2, LGR6, AZGP1 and SERPINA12 is depicted in Figure 4.7. Similarly to highly upregulated genes, there seems to be a gradient across histological subtypes, which warranted further investigation of this trend (discussed in greater detail in Chapter 4.3.3.).

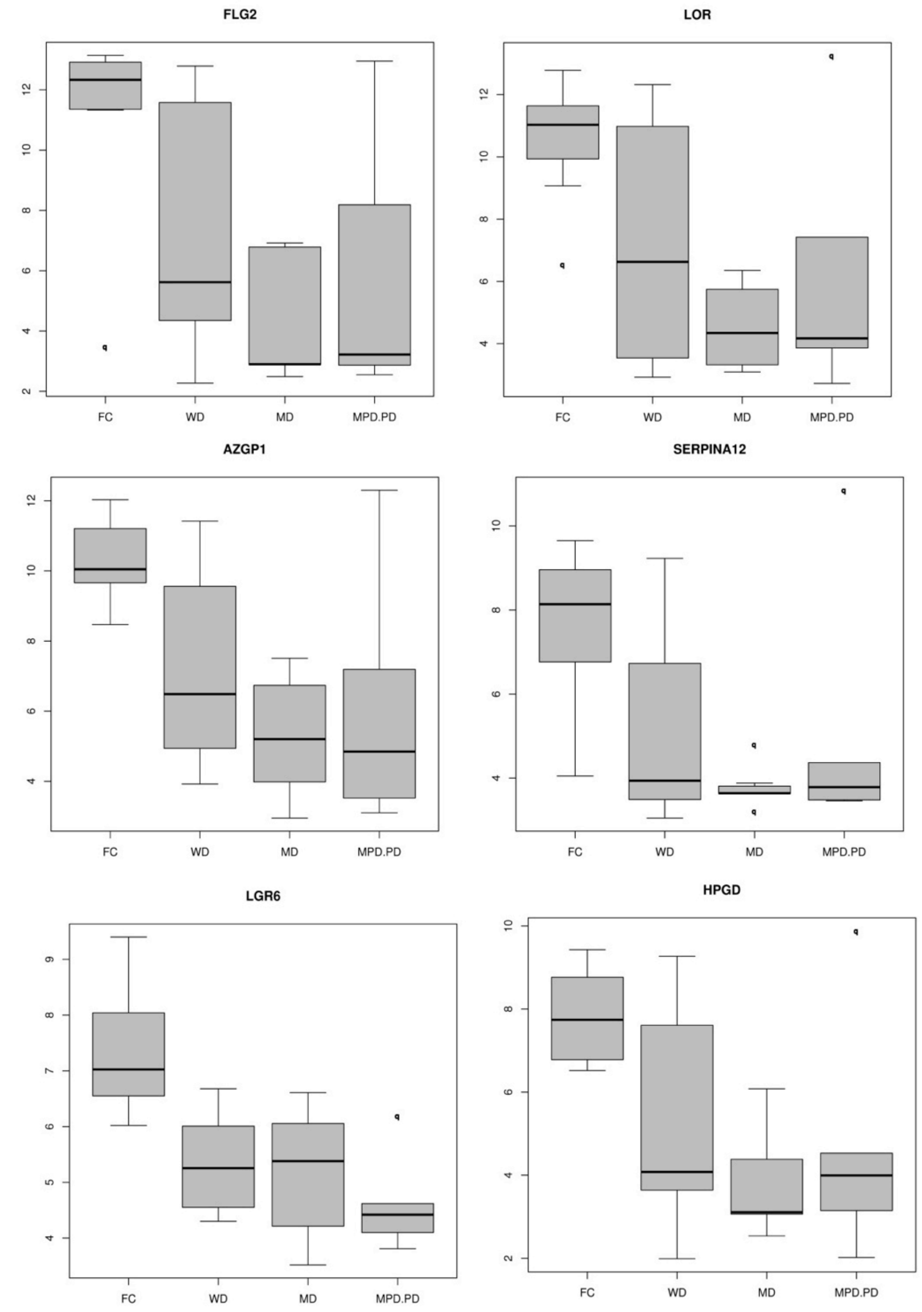


Figure 4.7 Average expression of selected down-regulated genes across different histological cSCC subtypes. A gradient from WD to less differentiated cSCC subtypes is apparent in the expression levels of FLG2, LOR, AZGP1 and LGR6. FC=AK.

4.3.2. Transcriptional profiling of matched AK and cSCC

A comparison between a total of 8 matched AK and cSCC pairs has revealed only 2 genes to be differentially expressed with the most stringent p value (<0.0001): MMP10 and PTHLH. Using a less stringent p value of < 0.01, 77 unique genes were shown to be differentially expressed between the two tissue types. 48 of these genes (62.5%) were listed among the 346 DEG discovered by global comparison of the two tissue types.

Shared genes included many of those listed among top 50 DEGs with the highest foldchange: ACVRA2A, ANXA1, CDH3, HN1, INHBA, ITM2A, KRT9, LAMC2, LRIG6, MET, metalloproteinases 1, 3, 10, PLAU, PLAUR.

4.3.3. Correlation of histopathological classification with expression profile

It was previously shown that genetic profiling of cSCC correlates with histological subtype (Purdie et al., 2009). I observed a similar trend in the cSCC transcriptome: as shown in Figure 4.8, which plots Pearson's correlation levels between individual tumour samples based on all microarray probes. A cluster of 11 WD tumours and 2 MD tumours is apparent on the right, a cluster of 3 MD, 1 MPD and 2 WD tumours is seen in the center, and a cluster of 1 WD, 2 MD, 1 MPD and 1 PD samples on the left, in addition to a cluster of 1 PD and 1 MPD sample on the left extreme.

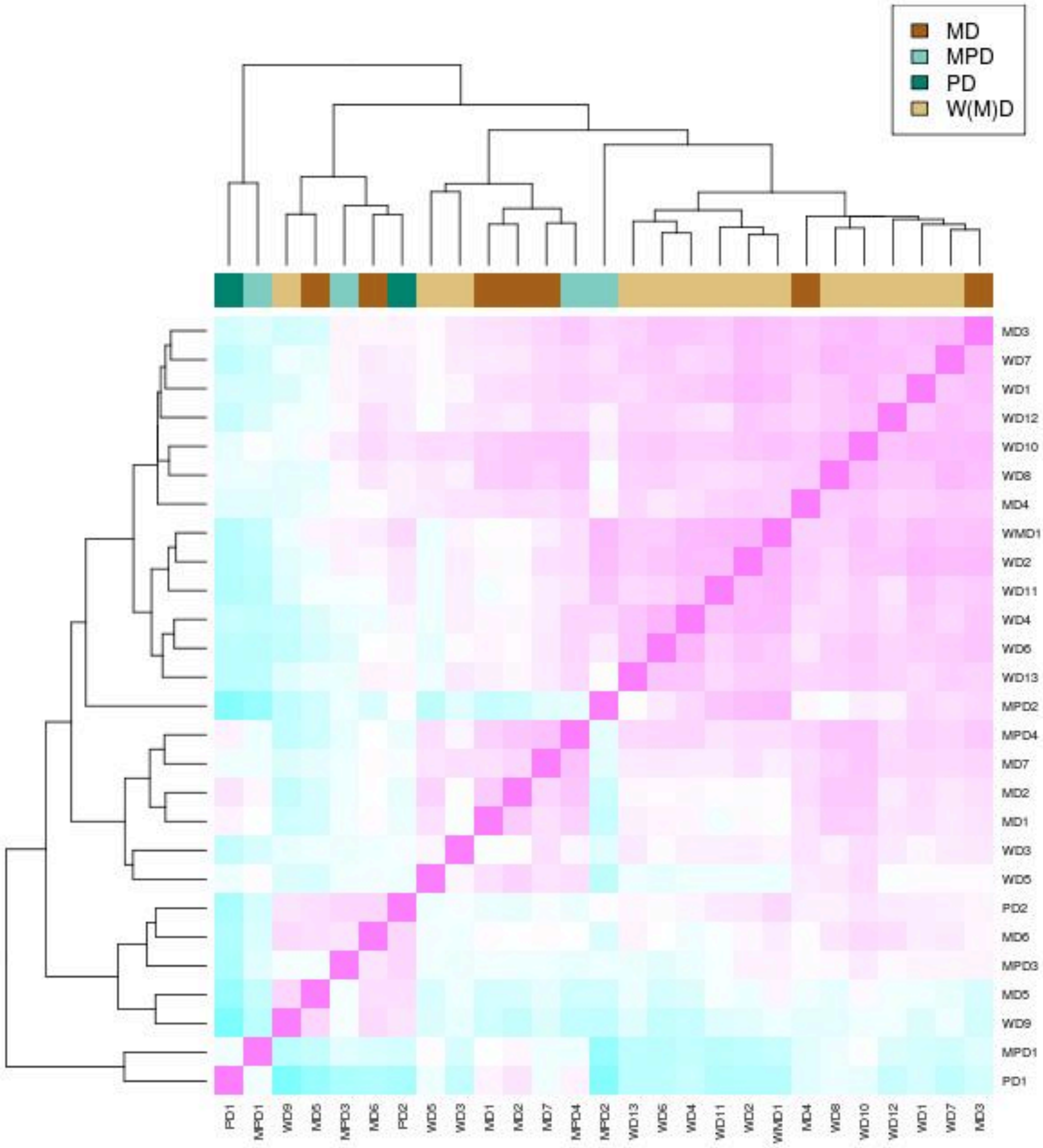


Figure 4.8 Unsupervised hierarchical clustering of different histological subtypes of cSCC. A cluster on the right contains predominantly WD cSCC, middle cluster contains a mix of histological subtypes, and a cluster on the left represents mostly MD/PD cSCC. For the purpose of this plot, all WD tumours and and 1 WMD cSCC were group into one: W(M)D.

I then explored the statistical significance of clusters using bootstrap correlation. As shown in Figure 4.9, 5 unequivocal clusters are present if the whole microarray values are subjected to bootstrap correlation. Except for 1 WD sample in a cluster of 5 higher-grade tumours, and a WMD sample clustering with WD samples, as expected, those clusters reflect the histological subtype of tumours.

ANOVA testing also showed that histopathological subtype affects the expression profile (p value < 0.0001), and this was confirmed by linear regression analysis, in which histological grade was the only variable to affect expression among those in the model (histological subtype, gender and immune status).

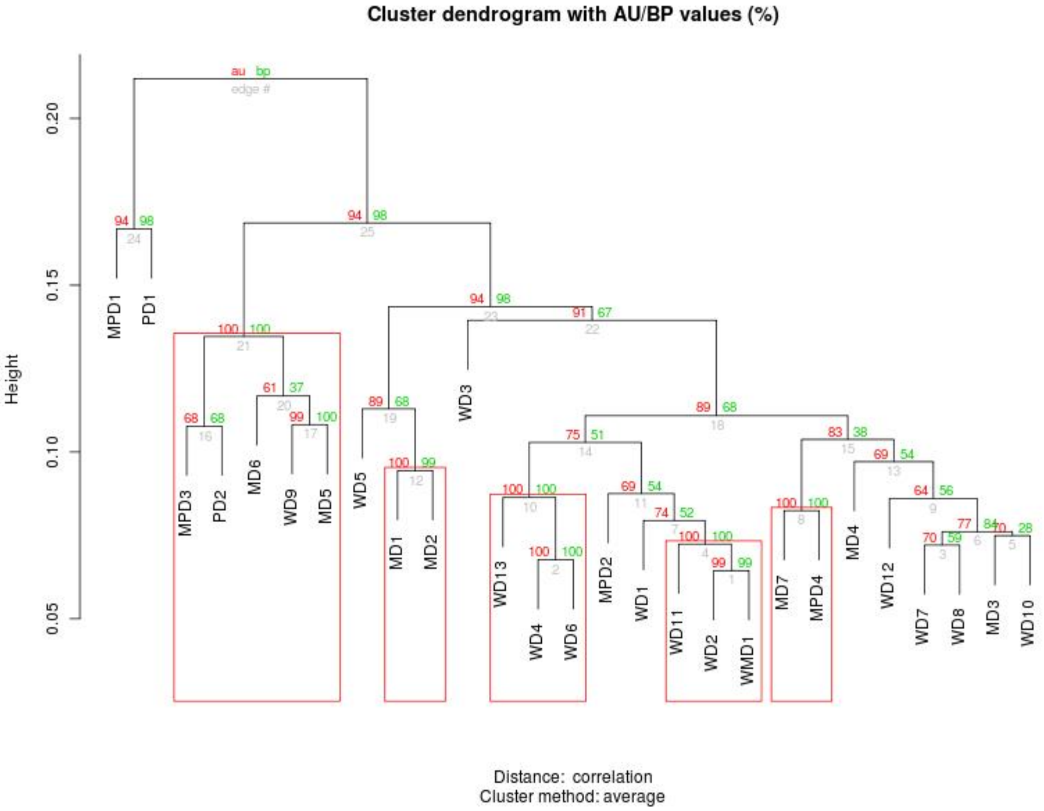


Figure 4.9 Bootstrap correlation of cSCC histological subtypes. Red frame indicates statistically significant clusters.

Plotting corresponding probes between 4 histologically distinct tumours has revealed a rather widespread distribution of expression values between a PD and MPD tumour, unlike in comparison between WMD and WD tumour (Figure 4.10). Based on this plot, I suspected that significant transcriptional differences may exist between individual histological subtypes. These are discussed in greater details in Chapter 4.3.3.1.

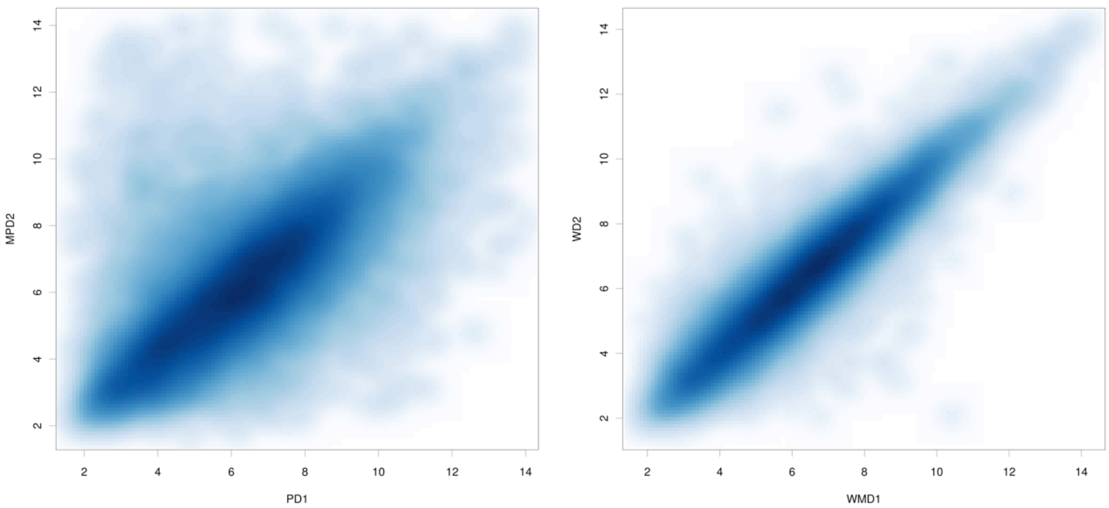


Figure 4.10 Log2 levels of corresponding probes in PD and MPD tumour (low correlation, left panel), and WMD and WD tumours (high correlation, right panel).

Additionally, I calculated Pearson's correlation between individual histological subtypes as the following: WD and MD; WD and MPD; WD and PD; MD and MPD; MD and PD; MPD and PD.

The correlation boxplot is shown in Figure 4.11 and clearly indicates that similarity between WD and higher tumour grades inversely correlates with decreasing tumour differentiation. On the other hand, the correlation between higher tumour grades is relatively constant.

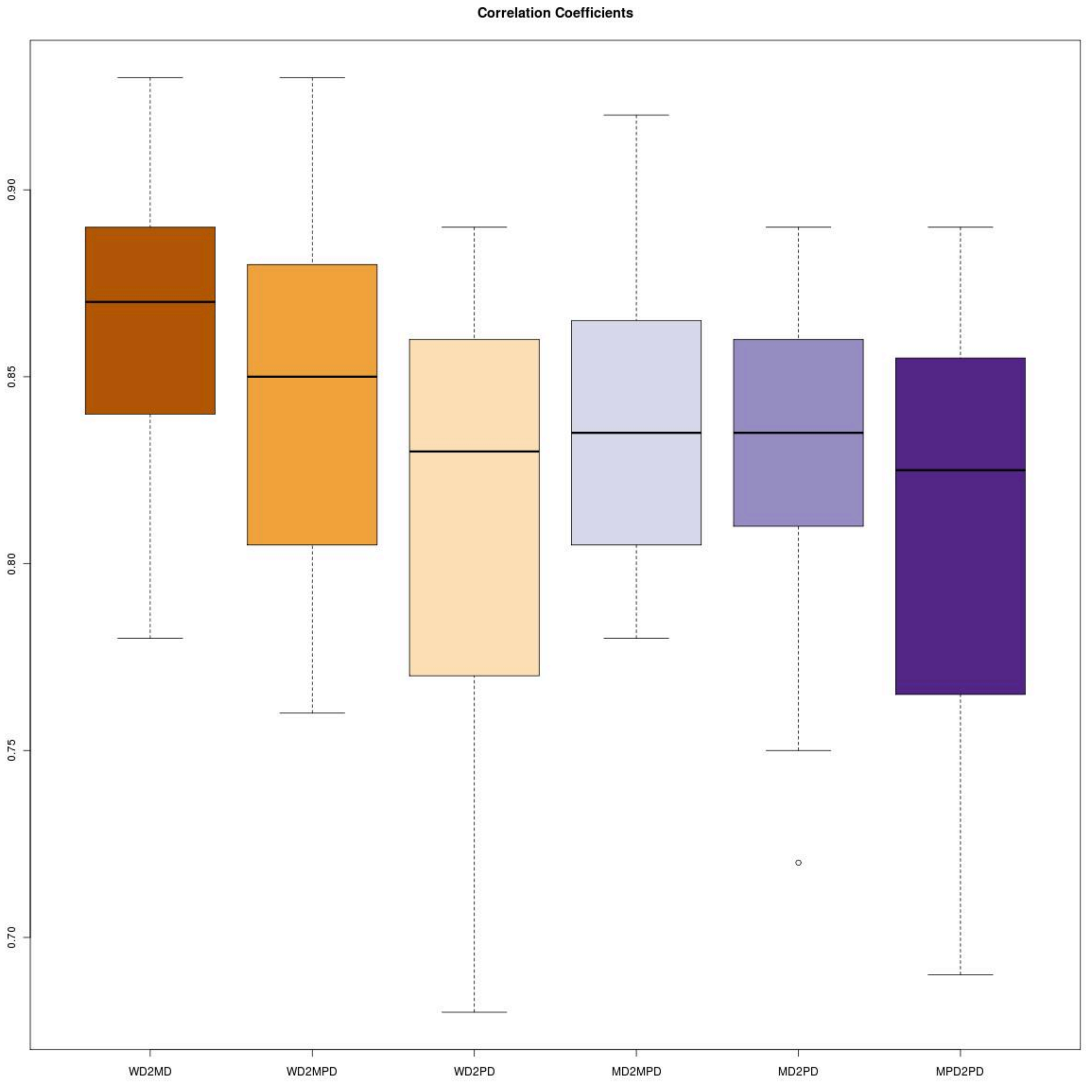


Figure 4.11 Correlation of array-wide log2 expression values between different cSCC histological subtypes. As tumour become less differentiated, the correlation level with WD cSCC decreases.

Due to the low number of PD tumours in the dataset, I then regrouped the MPD/PD samples with PD samples. As shown in Figure 4.12, correlation level is highest between identical histological cSCC types (WD and WD, MD and MD), and gradually decreases as less differentiated tumours are correlated with WD and MD samples. P values derived with Student's t-test assessing the mean correlation values between each tumour subtype are shown in Table 4.2. These results indicate that the similarity level between different histological subtypes is significantly different.

Subtype correlation	P-value (Student's T-test, unpaired)
WD|WD versus WD|MD	0.004
WD|WD versus WD|MD/MPD	<0.0001
MD|MD versus MD|MPD/PD	<0.0001

Table 4.2. P values asessing the hypothesis that correlation levels between tumours of the same and different histological typing are the same. These results demonstrate that the level of transcriptional similarity is significantly different between these groups.

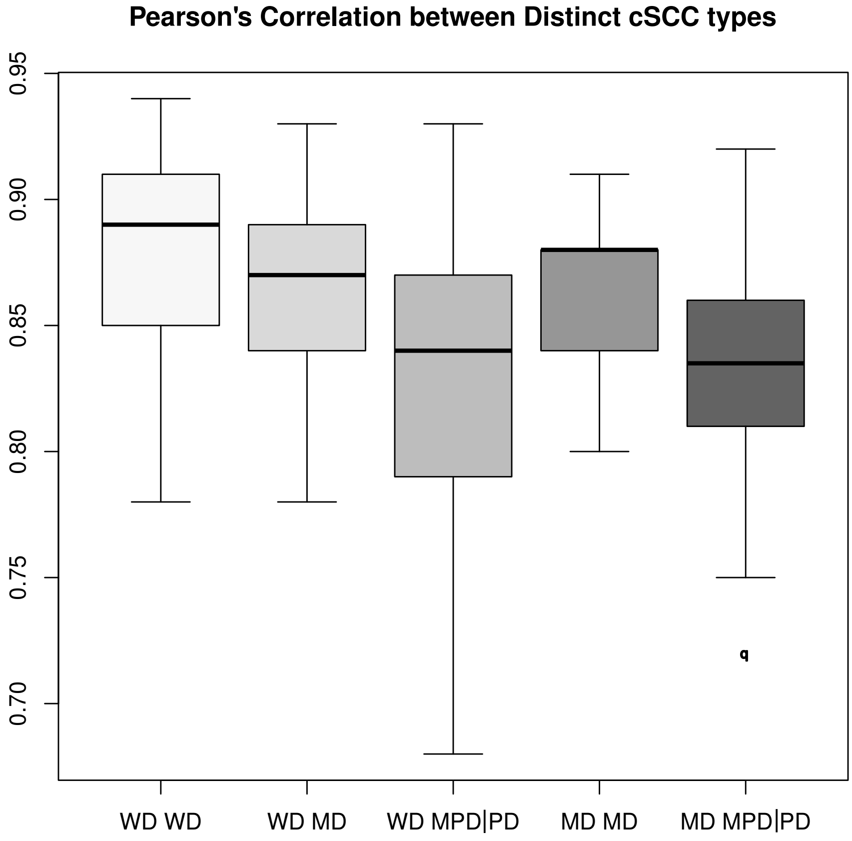


Figure 4.12 Pearson's correlation between distinct histological subtypes. MPD/PD and PD tumours were grouped together for this comparison. The level of similarity gradually decreases with the comparison of WD and MD tumours to less differentiated subtype(s).

4.3.3.1. Transcriptional differences between distinct histological subtypes of cSCC

Based on the results presented in previous chapter, I hypothesised there may be transcriptional differences in specific genes between the distinct tumour subtypes.

The comparison of WD and MD cSCC has not detected any significantly differentially expressed genes after adjusting the p value. A comparison between WD and MPD-PD and PD samples combined revealed a total of 18 probes that are differentially expressed with the most stringent p value (<0.0001, Table 4.3).

Symbol	Description	Chromosome	logFC	adj.P.Val
NA	NA	/	2.099086068	<0.0001
NA	NA	/	1.337295268	<0.0001
TOPORS	topoisomerase I binding, arginine/serine-rich	9	1.116112328	<0.0001

EPB41L5	erythrocyte membrane protein band 4.1 like 5	2	-1.347575898	<0.0001
NA	NA	/	-1.440668934	<0.0001
GRAMD2	GRAM domain containing 2	15	-1.872930679	<0.0001
FGFR2	fibroblast growth factor receptor 2	10	-1.977345253	<0.0001
FGFR2	fibroblast growth factor receptor 2	10	-2.012884607	<0.0001
EHF	ets homologous factor	11	-2.051027316	<0.0001
FGFR2	fibroblast growth factor receptor 2	10	-2.205722516	<0.0001
NA	NA	/	-2.227342478	<0.0001
EHF	ets homologous factor	11	-2.295503521	<0.0001
FGFR2	fibroblast growth factor receptor 2	10	-2.311526316	<0.0001
EHF	ets homologous factor	11	-2.345541049	<0.0001
EHF	ets homologous factor	11	-2.426611372	<0.0001
CNGB1	cyclic nucleotide gated channel beta 1	16	-2.924659532	<0.0001
PTPRZ1	protein tyrosine phosphatase, receptor-type, Z polypeptide 1	7	-3.744986849	<0.0001

Table 4.3. Differentially expressed probes between well-differentiated tumours and moderately-to-poorly and poorly-differentiated tumours. LogFC reflects direction of change from WD to higher grade *tumours.*

This probeset has revealed a total of 7 genes to be differentially expressed between the tumour subtypes. CNGB1 was previously shown to be up-regulated in metastatic oral SCC (Mendez et al., 2007). Protein tyrosine phosphatases are known to play a role in cancer (Motiwala and Jacob, 2006), and PTPRZ1 is down-regulated by ZEB1 (Rokavec et al., 2012), a known epithelial-to-mesenchymal transition (EMT) marker, and its loss has been linked to prostate cancer metastases (Diamantopoulou et al., 2012). EPB41L5 has been previously linked to epithelial-to-mesenchymal transition regulation (EMT) (Hirano et al., 2008). The only upregulated gene in the set, TOPORS, is known to bind to p53 and acts as a tumour suppressor (Yang et al., 2010). The loss of EHF expression has been shown in various invasive cancers, including bladder, oral, prostate and breast cancer (Mathsyaraja and Ostrowski, 2012). FGFR2 is an established oncogene, involved in MAPK, Ras, PI3K-Akt pathway, yet the loss of its signalling has been linked to EMT initiation (Diamantopoulou et al., 2012). In our dataset, FGFR2 levels gradually increased during progression from skin through AK to higher grade tumours (Figure 4.13).

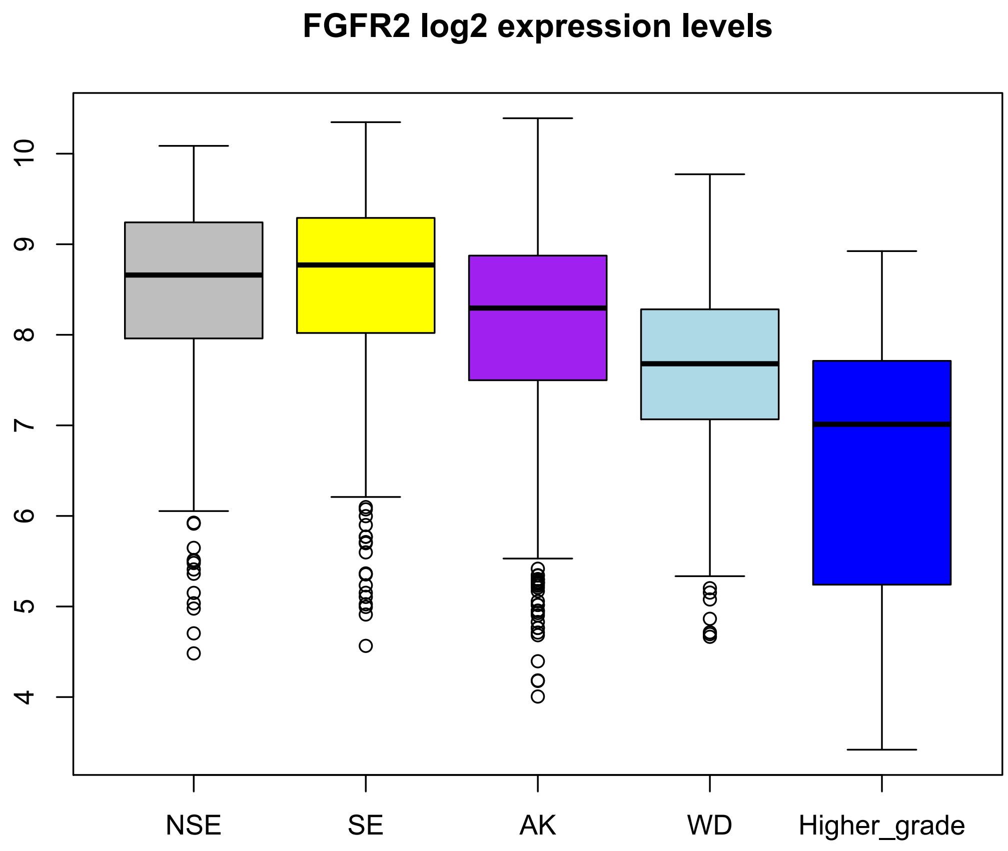


Figure 4.13. FGFR2 log2 expression levels. Black line=median. Higher_grade = MD, MPD or PD tumours. WD=well-differentiated. This figure shows gradual progression of FGFR2 expression loss during the oncogenic transformation of skin to cSCC. This was a statistically significant observation (p value<0.0001, ANOVA). Empty circles=outliers.

4.3.4. Transcriptional analysis of genes relevant in stem cells

4.3.4.1. Genes upregulated in Epidermal Stem cells

I then hypothesised that various epidermal stem cell markers may be differentially expressed in cSCC compared to AK due to the dedifferentiation of malignant keratinocytes and potential acquisition of stem cell-like properties. This phenomenon has been previously described in intestinal tumourigenesis (Sauermann et al., 2008).

Several genes that were previously shown to be involved in epidermal stem cells were present on the microarray. Of these, two upregulated genes were found within our top 508 probes: MAP4, a microtubule regulator involved in cell cycle progression, and TCF3, additionally shown to be indispensable for skin homeostasis (Nguyen et al., 2009). Other upregulated genes related to epithelial cancer stem cells included TNC (tenascin, $p = 0.0001$), PDPN (podoplanin, $p < 0.001$), ASXL1 and RUNX1 (both p < 0.01), SOX9 (p = 0.07). The full list included MAP4, LRIG1, FRMD4A, PHKA2, ASXL1, C20orf111, RUNX1, TNC, SOX9, GAS6, TCF3, BMI1 and PDPN.

When all probes for these genes present on the array were plotted as a heatmap, there was no visible trend in hierarchical clustering, and the majority of genes were expressed at rather low or mid-range levels. Interestingly, replicate probes for identical genes did not cluster together, and some detected expression levels at the opposite spectrum of intensity (PDPN, TCF3, Figure 4.14). This is likely due to the presence of different transcripts of those genes.

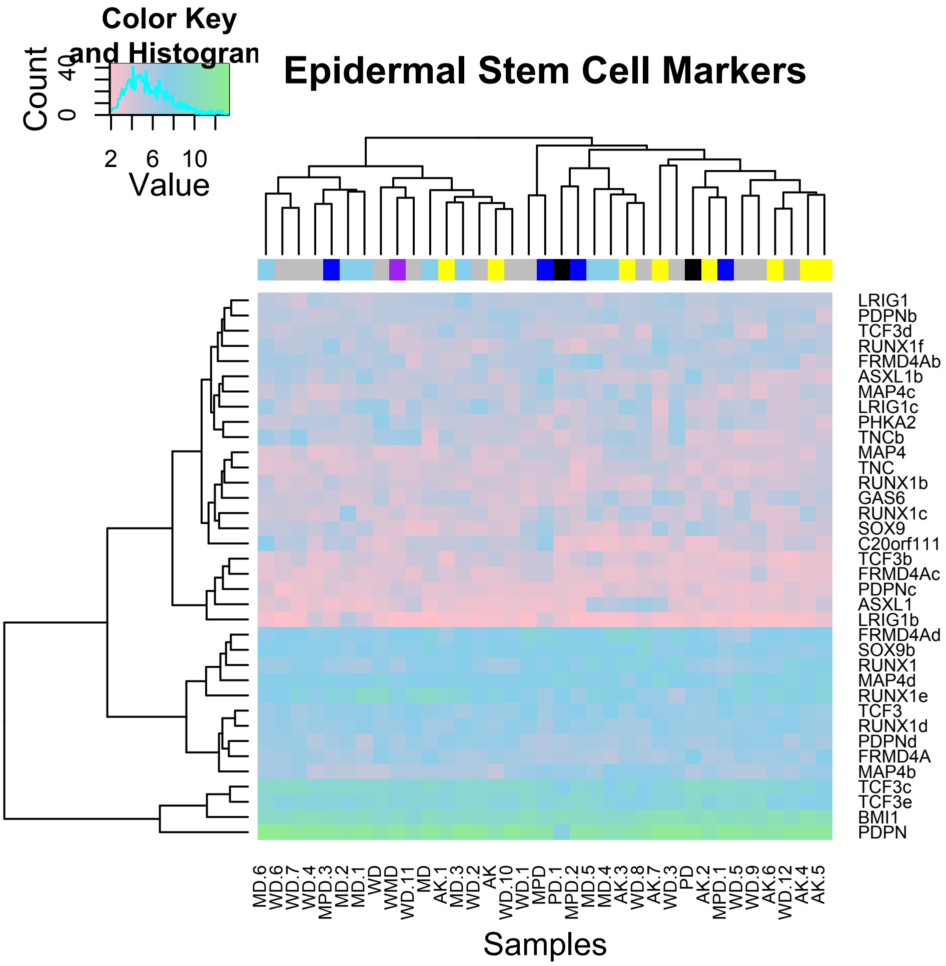


Figure 4.14 Heatmap of log-transformed expression levels of epidermal stem cell markers in AK and cSCC. Yellow=AK, black=PD, grey=WD, purple=WMD, light blue=MD, dark blue=MPD. This set of markers is not sufficiently differentially expressed across the two tissues to segregate those based on unsupervised hierarchical clustering.

I hypothesised that because the oncogenic process has already begun in AK, the transcriptional differences in epidermal stem cell markers are more pronounced in the transition from normal skin to AK and cSCC. Thus I plotted the full set of transcription data, including expression profiling data from normal skin and RNALater treated AKs described in the previous chapter. The probeset was able to almost completely separate the three tissue types and create a certain gradient of expression levels between skin, AK and cSCC (Figure 4.15). Additionally, most probes for identical genes clustered together.

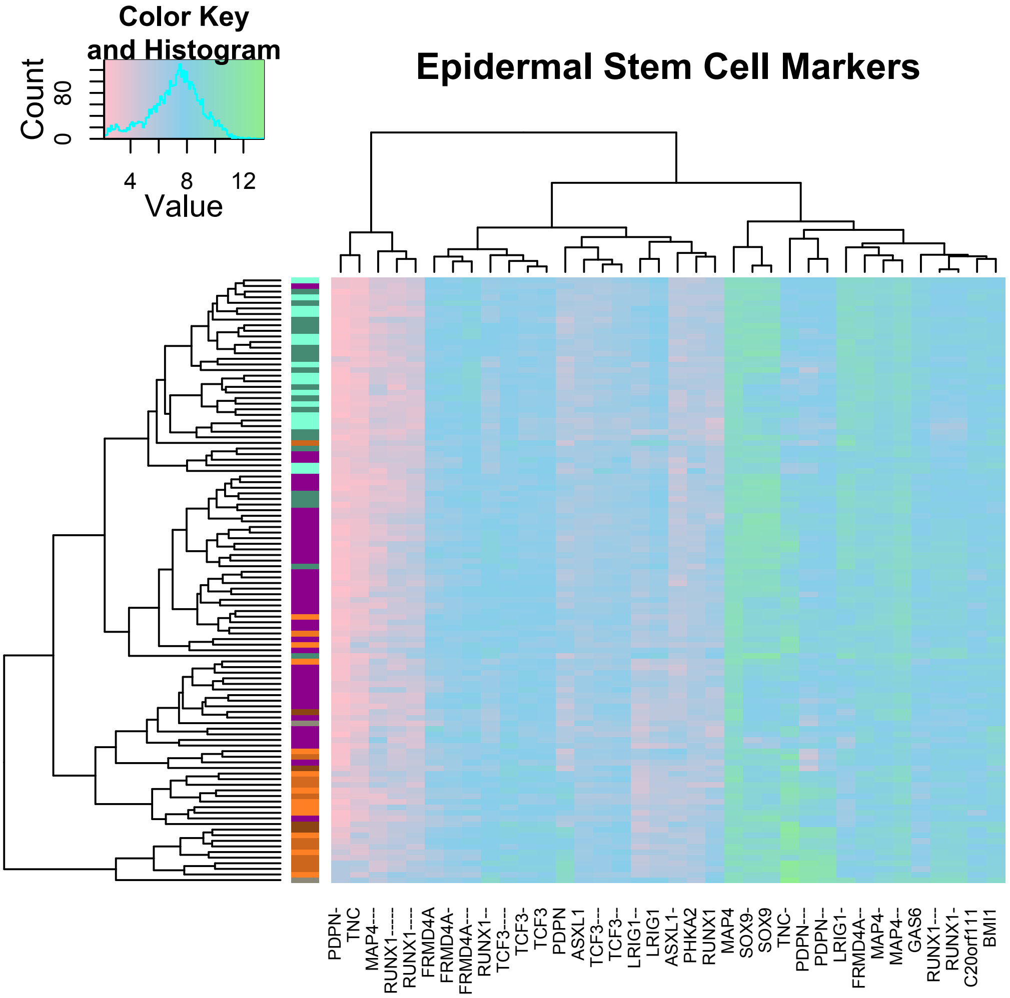


Figure 4.15 Epidermal stem and cancer stem cell markers plotted across all study samples. The majority of skin, AK and cSCC samples formed distinct clusters based on this probeset. Light green=NSE, dark green=SE, purple=AK, increasing tones of brown = increasing cSCC grade.

To verify that the observed hierarchical clustering is not due to inherently skewed transcription profiles of the tissues and to exclude the posibility that any set of 36 genes would lead to identical clustering, I plotted randomly selected 36 probes. Unsupervised hierarchical clustering was less homogeneous, and previously observed gradient was absent (Figure 4.16).

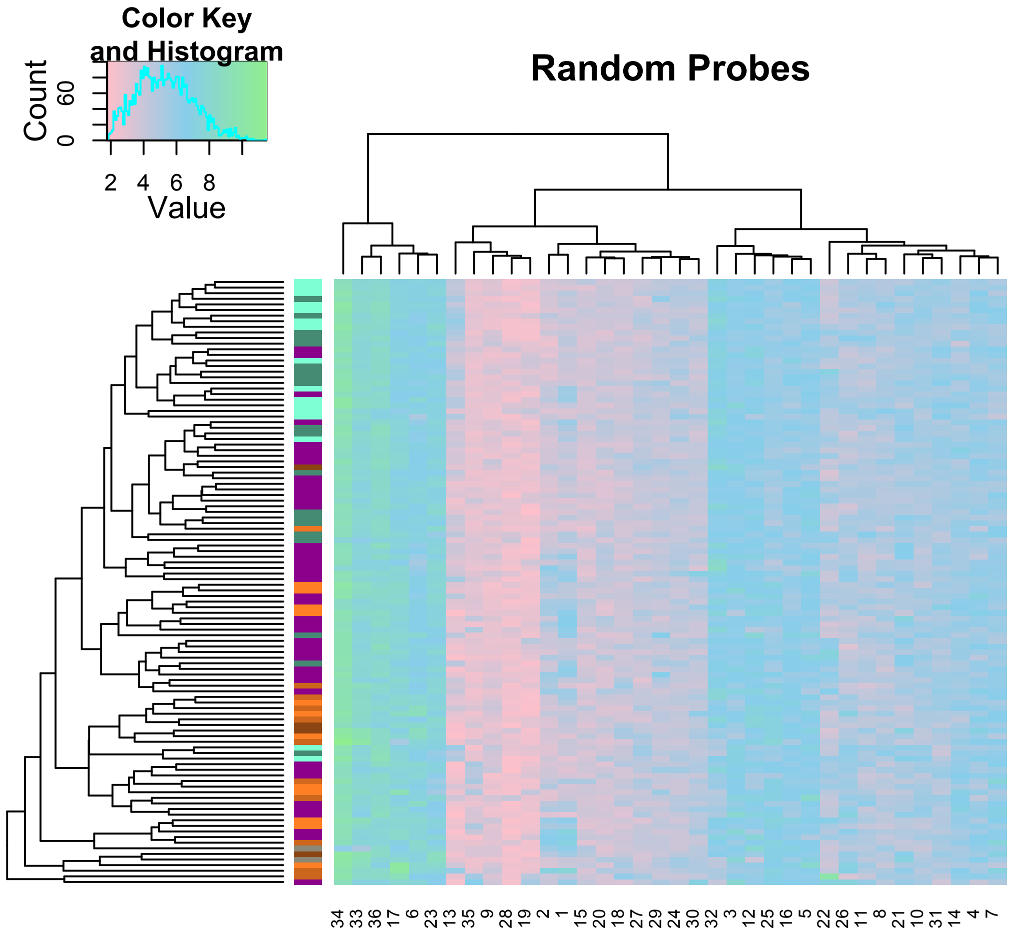


Figure 4.16 Random set of 36 probes plotted across all samples. Light green=NSE, dark green=SE, purple=AK, increasing tones of brown = increasing cSCC grade.

To explore the statistical significance of this observation, I used bootstrap correlation for both probesets to calculate if any of the clusters observed are statistically significant: using epidermal stem cells marker, I detected a total of 4 significant tumour clusters, 1 AK and 2 normal skin clusters. Using random markers, I detected no statistically significant clusters.

I then further reduced the probeset to include probes with the most apparent gradient. Heatmap of this reduced probset has shown that tenascin and podoplanin are much more highly expressed in cSCC tumours compared to skin (Figure 4.17).

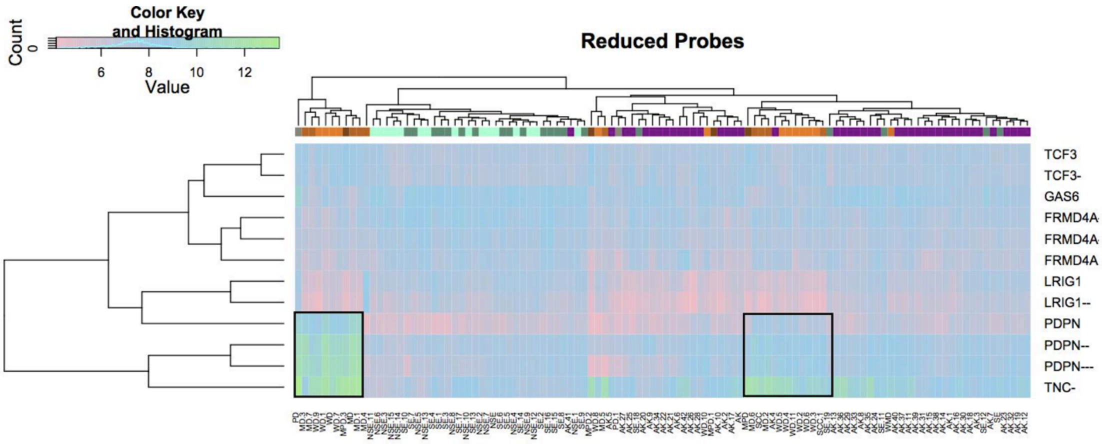


Figure 4.17 Heatmap of reduced number of probes with the most prominent expression level gradient. Boxes highlight podoplanin and tenascin expression levels in tumours. Light green=NSE, dark green=SE, purple=AK, increasing tones of brown = increasing cSCC grade.

This finding is consistent with previous reports that have shown that podoplanin is expressed in cSCC, and that its levels increase with increasing tumour grade (Schacht et al., 2005). Our results only partially confirm this trend (Figure 4.18), because the levels of PDPN suddenly drop in MPD and PD tumours (ANOVA comparing PDPN levels in WD, MD, MPD and PD tumours was not statistically significant, p value=0.29).

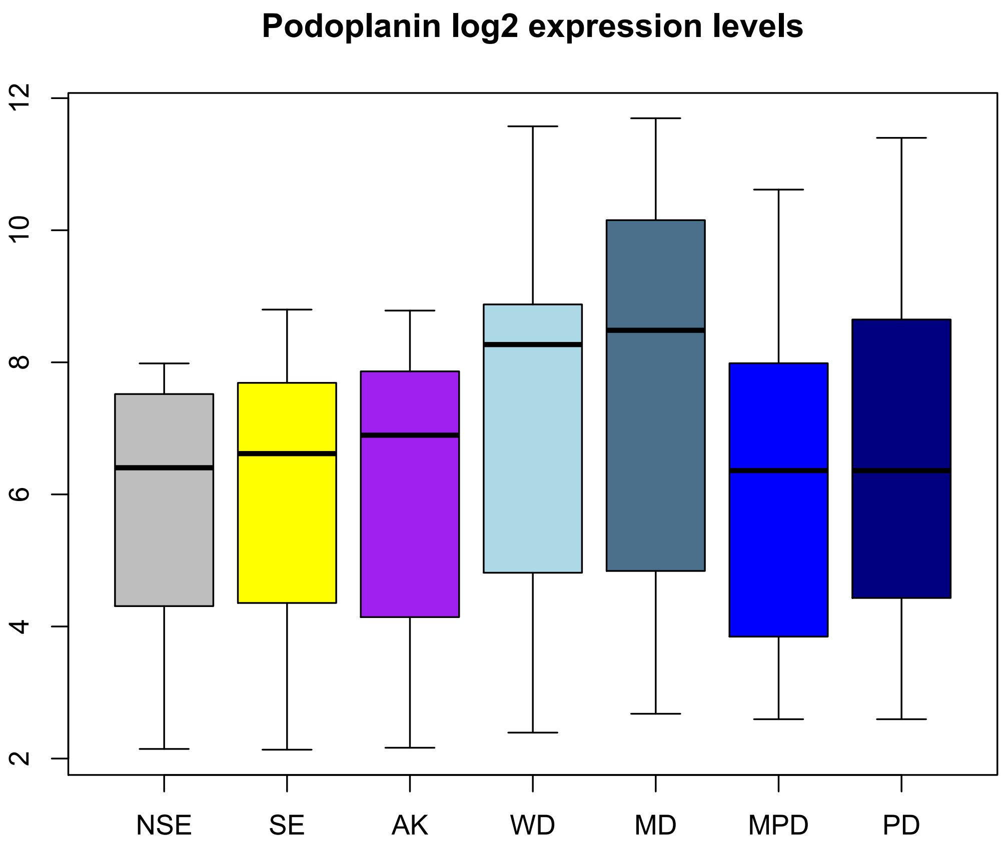


Figure 4.18 Podpolanin log2 expression levels across tissue groups. Black line indicates median. While PDPN levels gradually increase towards MD tumours, their levels then drop in MPD and PD samples. Expression level was stastistically significantly different comparing tumours to AK and cSCC (p value <0.0001), but not between different histological tumour subtypes.

Tenascin C (TNC) has been shown to promote EMT in colon cancer (Takahashi et al., 2013) and its increasing levels were previously linked with increased level of dysplasia in AK (Lentini et al., 2006). As shown in Figure 4.19, TNC levels increase with progression from skin to cSCC, suggesting that the process of EMT may contribute to the malignant process.

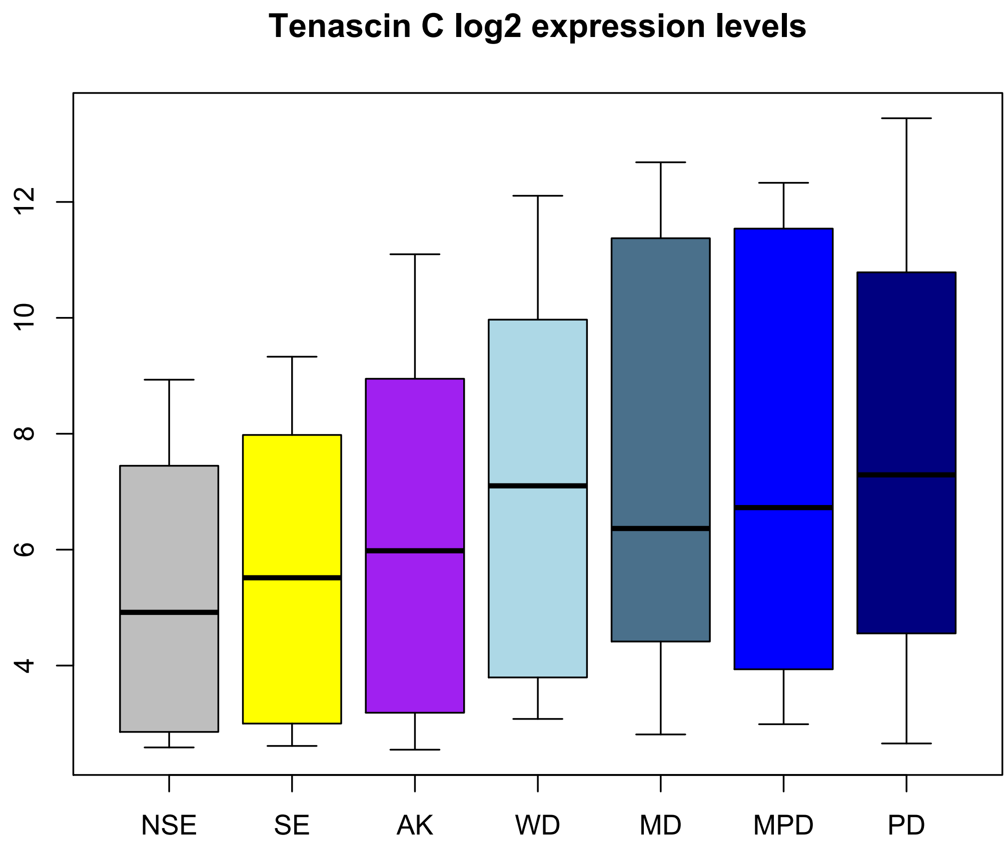


Figure 4.19 Log2 expression levels of Tenascin C across our sample series. The median levels presented by black line in the boxplot are higher in tumours compared to normal skin and AK (p value=0.029, ANOVA). The difference in expression levels between tumour subtypes was not statistically significant (p value=0.9).

Figure 4.17 also shows that LRIG1 levels were decreased in cSCC and a subset of AK compared to normal skin. Given that LRIG1 has been shown to represent a marker of epidermal stem cell multipotency in mouse skin (Jensen et al., 2009), this observation was contrary to my hypothesis that markers of stemcellness increase their expression during the progression from skin to cSCC. However, LRIG1 has been previously shown to be decreased in cSCC (Tanemura et al., 2005), and its levels inversely correlated with tumour differentiation and poor prognosis. Additionally, it has been suspected to be a tumour suppressor gene (Hedman et al., 2002), thus this observation is consistent with previous findings.

4.3.4.2. Other stem cells markers lead to clustering of AKs and WD cSCC

Many markers described as being related to stem cells (other than epidermal cancer stem cells) were also present on the array. Some of those were within our top geneset (CD 47, marker of leukaemia stem cells, while some such as HAS2, umbilical cord stem cells marker, showed non-significant upregulation (p = 0.47).

This extended probeset of stem cell markers included CD44, EPCAM, ABCG2, BMI1, CD29, Nestin A, Nestin B, STAT3, ABCB1, NOTCH1, NOTCH2, NOTCH3, THY1, WNT2b, CD47, CD59, CD24, CD99, CD151, CD209, CD84, CD200, CD14, CD34, CDCP1, HAS2, KRT15, KRT19, LRG5, LRP4, MCM2, NRP1, PCNA, PTPRC, SOX10, SOX2, SOX4, SOX6, SOX7, SOX6 and SOX9.

When applied to AK and cSCC samples, this probeset lead to clustering of 5/8 AK samples, and 10/14 WD/WMD samples (Figure 4.20).

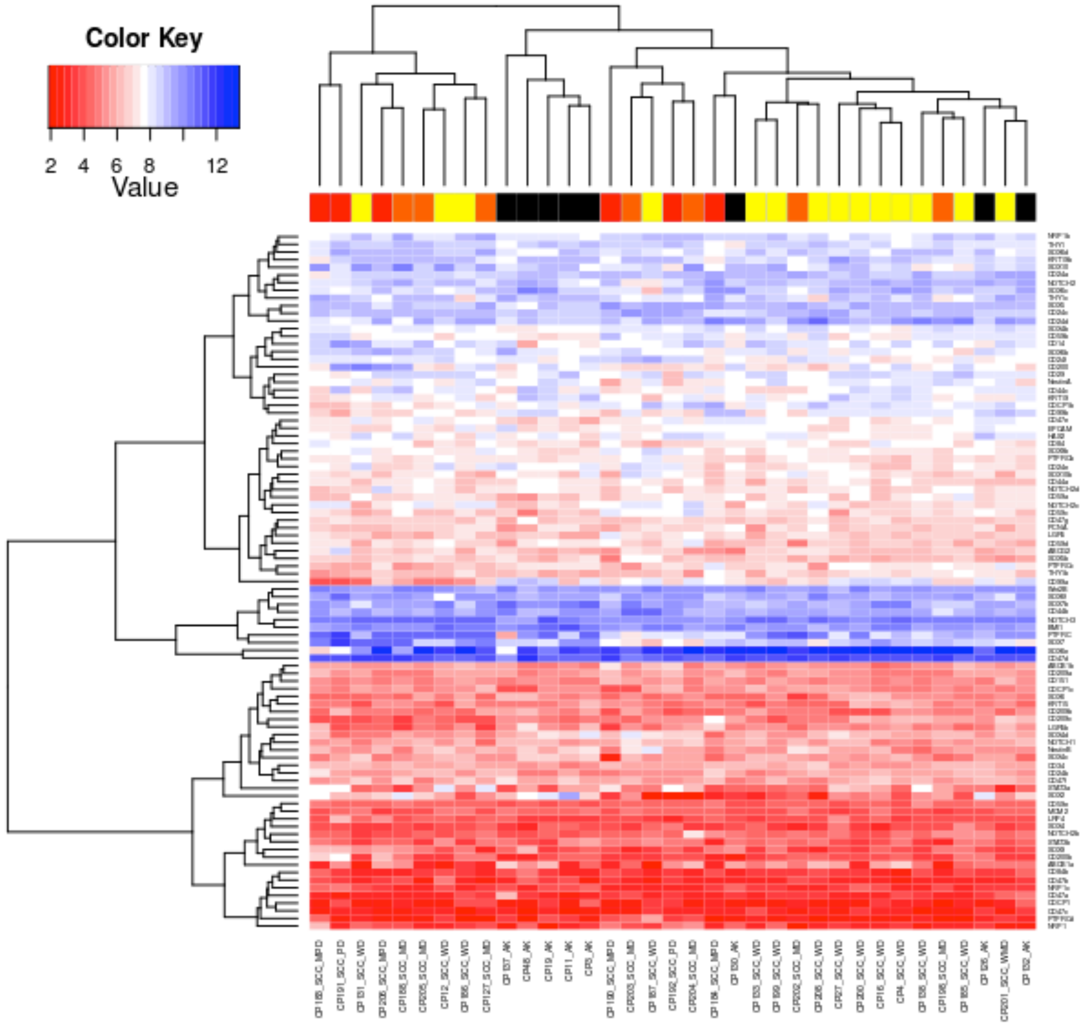


Figure 4.20 Log2 expression levels of stem cell markers. Black=AK, yellow=WD/WMD, orange=MD/MPD, red=PD. This probeset generated a cluster of 5 AK samples, and 10 WD samples.

I then reapplied the same probeset to tumour samples in order to observe clustering based on differentiation status to confirm the trend of WD tumours to form a cluster. I expected that well-differentiated tumours would cluster together, and higher grade tumours would form independent clusters based on these markers. As shown in Figure 4.21, three clusters formed: One with 10/14 WD and 2/7 MD, one with 2 MD, 2 MPD-PD and 1 WD samples, and one with 3 WD, 3 MPD-PD and 3 MD samples.

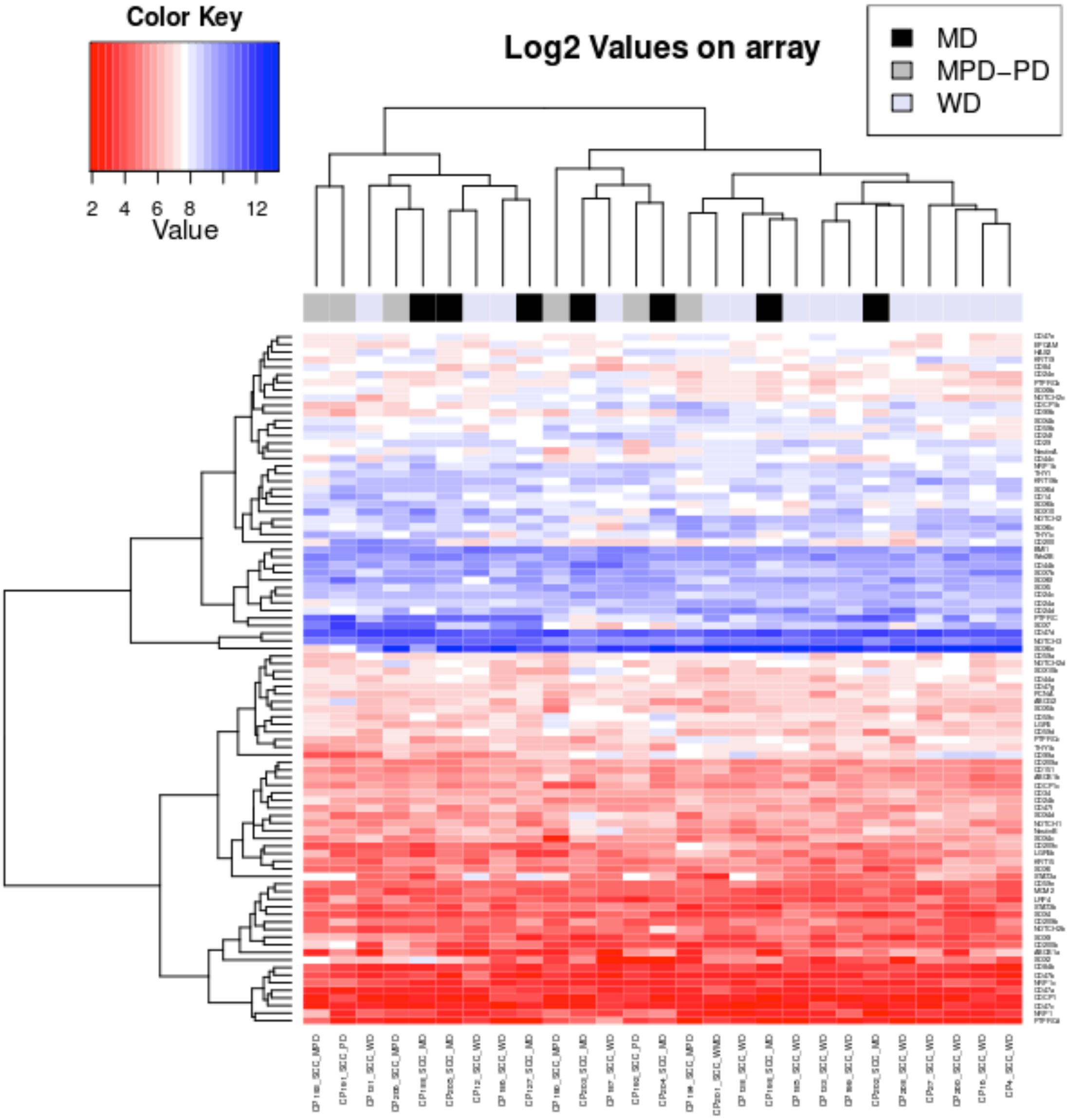


Figure 4.21 Hierarchical clustering of distinct histological cSCC subtypes based on selected stem cell markers. 10/14 WD cSCC form a cluster on the right, 10/13 of less-than-well differentiated cSCC form the two remaining clusters (chi square 0.021).

These data suggests that – although not in quite a linear fashion by increasing from WD through MD to PD tumours – differences in stem cell markers expression herald progression from AK towards higher grade malignancies.

Because this clustering trend was similar to that based on overall transcription profile (Figure 4.8), I proceeded to verify that this trend would not occur if any set of related probes was used for unsupervised clustering. I used a set of 42 probes involved in Ras signalling, and as shown in Figure 4.22, this probeset has not generated the same clusters.

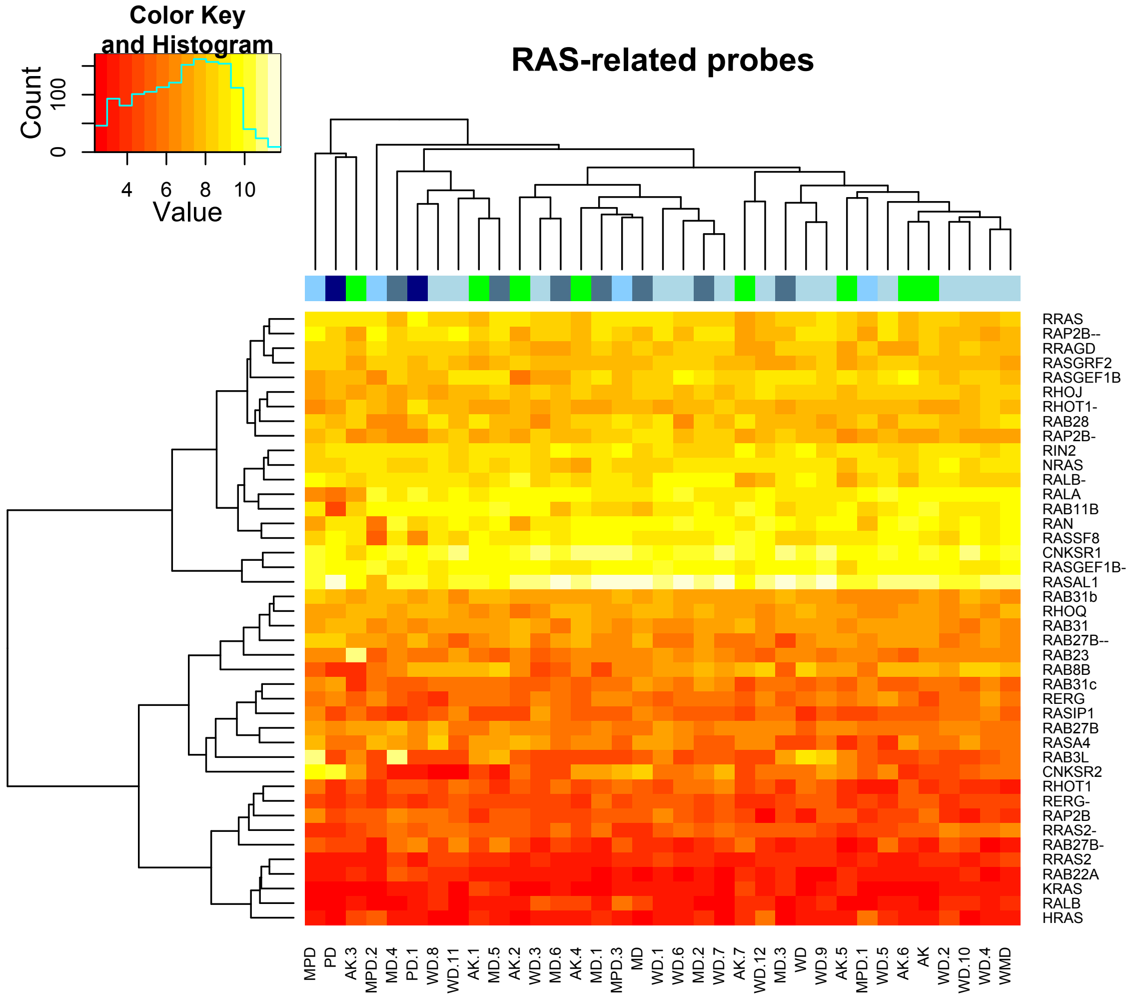


Figure 4.22 Unsupervised hierarchical clustering based on probes relevant in Ras signal transduction. 8/14 WD tumours clusterin the top-right cluster, 10/15 cluster in the remaining gour clusters (p value=0.09, chi square). Green=AK, light blue=WD, grey blue=MD, sky blue=MPD, navy blue =PD.

4.3.4.3. Genes associated with epithelial-to-mesenchymal transition

Several genes associated with EMT were present on the array: ACTA2, ACTG2, CDH1, CDH11, CDH19, CDH22, CDH24, CDH26, CDH3, CDH5, CDH6, CLDN1, CLDN11, CLDN12, CLDN5, CLDN7, CLDN8, CTNNB1, CTNND1, FGFR2, FN1, KRT18, KRT8, KRT9, MUC1, OCLN, SNAI2, TCF3, THBS1, THBS2, THBS3, TWIST1, TWIST2, VIM, ZEB1, ZEB2. Loss of keratin 8 and 18 represent a hallmark of EMT (Lee et al., 1993). Additionally, loss of E-cadherin represents a fundamental element of EMT, and SNAI2, ZEB1, ZEB2 repress its transcription by binding to E-box sequence in its promoter, while TWIST1 and TWIST2 repress E-cadherin promoter by recruiting the Mi2/NuRD/MTA2 complex (Lee et al., 2002).

As shown in Figure 4.23, this probeset is able to distinguish between skin, AK and cSCC, and generates statistically significant clusters of cSCC, AK and normal skin based on bootstrap correlation, suggesting that EMT markers may contribute to malignant progression in cSCC.

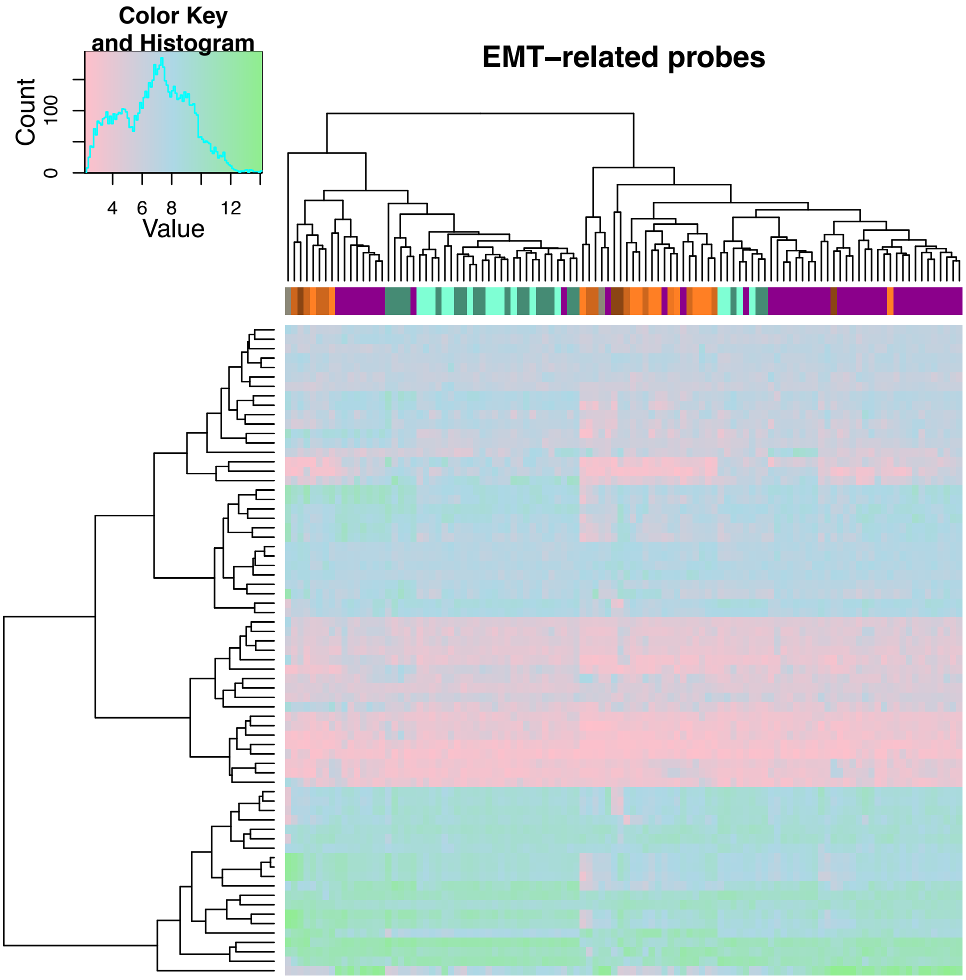


Figure 4.23 Unsupervised clustering of skin, AK and cSCC samples based on EMT probes. Skin, AK and cSCC form distinct tumours based on this probeset. Light green=NSE, dark green=SE, purple=AK, increasing tones of brown = increasing cSCC grade.

To explore this hypothesis further, I then looked at the mean levels of EMT markers in AK and cSCC subtypes (full list with mean expression values is provided in Appendix 5). E-cadherin (CDH1) down-regulation is a hallmark of EMT (Kang and Massague, 2004), and its expression level gradually decreases with increasing tumour grade in our dataset (Figure 4.24). Other important features of EMT such as decreased expression of claudins and occludins are also present in our data; claudin 11 (CLDN11) and cadherin 19, type 2 (CDH19) are differentially expressed between normal skin and AK and between AK and cSCC, while FGFR2 was differentially expressed between AK and cSCC only. Although occludin (OCLN) has not appeared in skin vs. AK or AK vs. cSCC DEG genelist, mean expression levels between normal skin and cSCC are statistically significantly different (p value <0.0001, t-test).

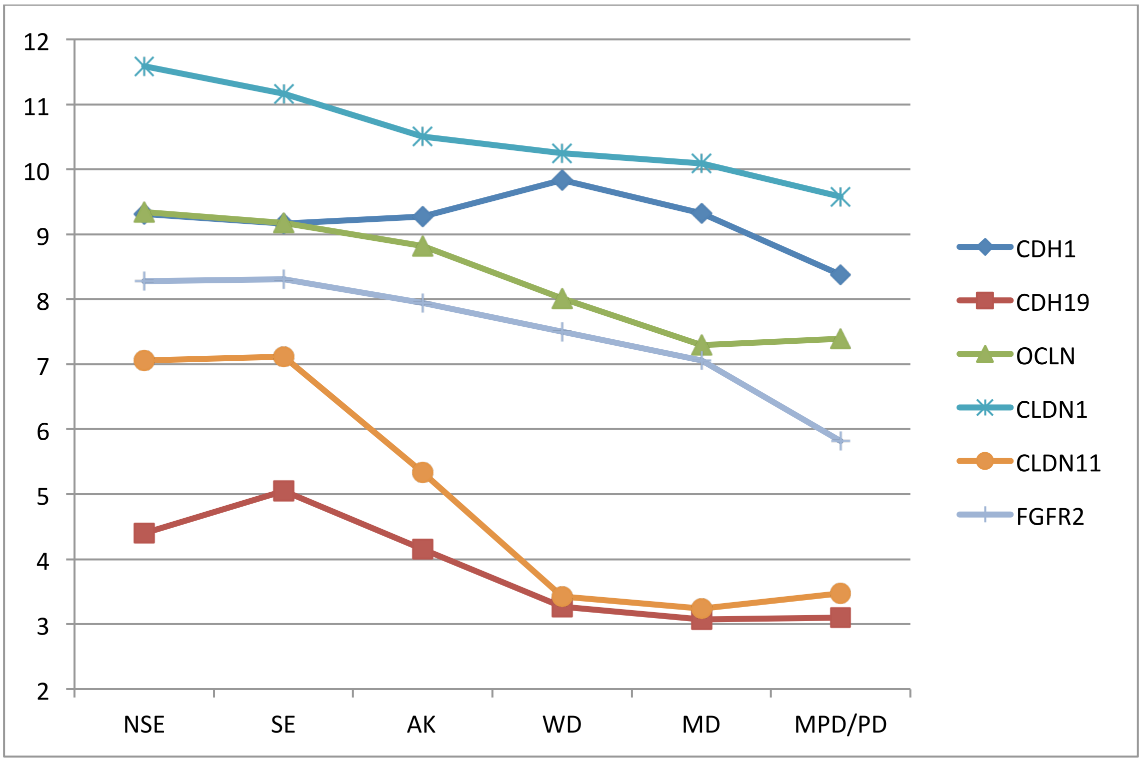


Figure 4.24 EMT markers that show decreasing trend across sample series. Loss or decreased expression of these markers signifies ongoing EMT. One-way ANOVA results (using log2 transformed expression values of NSE, SE, AK, WD, MD, MPD/PD for each gene as groups) showed statistically significant differences for CDH1, CDH19, OCLN, FGFR2, CLDN1 and CLDN11.

SNAI2 (slug) is a key regulator of EMT, and its mean expression level increases during the progression from skin through AK to cSCC (Figure 1.25). Although the difference in expression levels between normal skin and AK or AK and cSCC was not significant, ANOVA testing between skin, WD and MD, MPD/PD-PD tumours has revealed significant difference in expression levels (p value <0.0001).

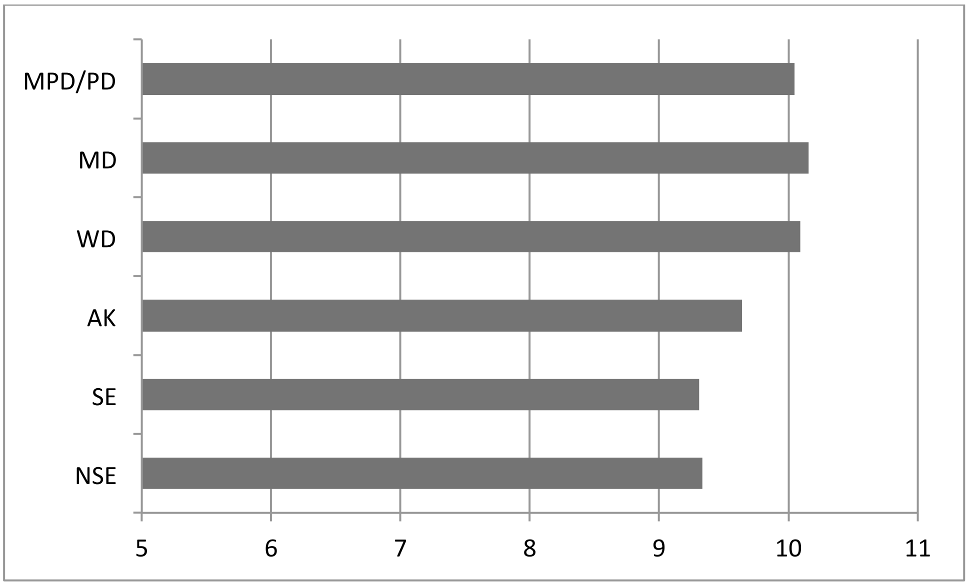


Figure 4.25 SNAI2 (Slug) mean expression level across our sample series. X axis=log2 transformed expression values. The mean expression level gradually increases during the progression of skin to cSCC.

TWIST1, MUC1 and ZEB2, all important EMT markers, increased their mean expression levels as tumour grade increased, yet ANOVA test has revealed that this is not a statistically significant observation (p value >0.05).

4.3.5. QPCR Validation

To validate the findings of the microarray analysis, thirteen genes were confirmed as either up- or down-regulated by quantitative real-time (qRT) PCR using an additional set of cSCC samples. These included both downregulated genes (ACVR2A and ID4) and upregulated genes (ADAM10, ADAM17, AKT1, ANXA1, INHBA, JUN, MET, MMP10, PAK2, PLK and PTHLH). In all cases, qRT-PCR showed statistically significant up- or downregulation between normal skin and cSCC, which was consistent with the microarray findings, suggesting a high level of reliability for the genes identified by the array. In particular, oncogenes JUN, MET and PAK2 were all confirmed to be overexpressed in cSCC.

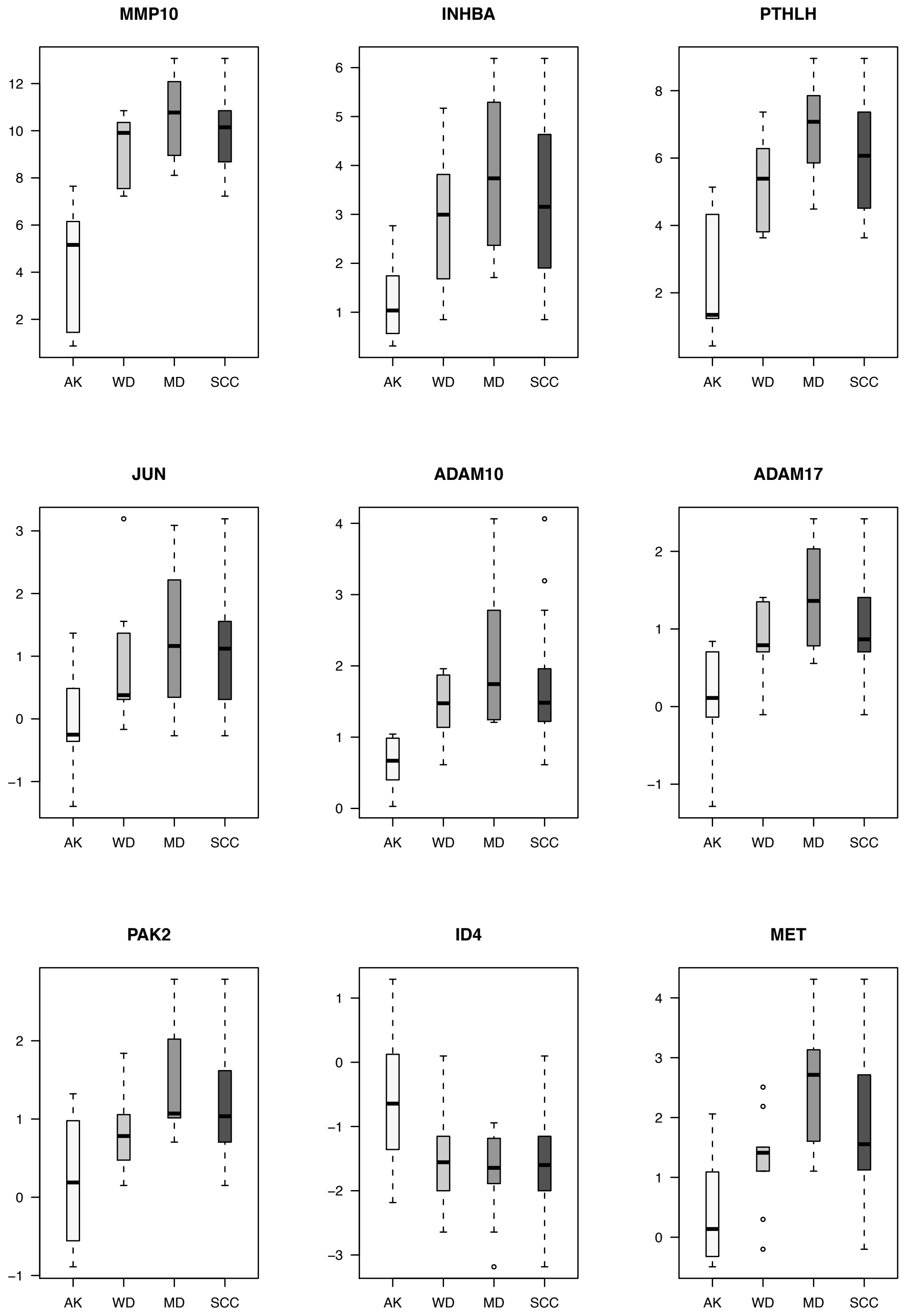


Figure 4.26 Differential expression of selected genes between AK and cSCC with QPCR. Y axis = normalised relative expression. Bold line=median. WD=well differentiated, MD=moderately differntiated cSCC, SCC=combined WD and MD values.

4.3.6. Gene Enrichment Analysis

4.3.6.1. Pathway analysis

KEGG Pathway analysis of the 508 differentially expressed probes showed the enrichment of 36 pathways (Table 4.4). Among the enriched pathways were several putative oncogenic pathways including MAPK, ErbB signalling pathways, TGF-beta pathway, several metabolic and biosynthesis-related pathways, and pathways involved in cell-cell interactions such as adherens junctions, focal adhesion and tight junctions.

KEGG Pathway ID	Pval	Adj.Pval(FDR)
Epithelial cell signaling in Helicobacter pylori infection	<0.0001	<0.0001
Carbon fixation in photosynthetic organisms	<0.0001	<0.0001
Cytokine-cytokine receptor interaction	<0.0001	<0.0001
Glycolysis / Gluconeogenesis	<0.0001	<0.0001
Adherens junction	<0.0001	<0.0001
Focal adhesion	<0.0001	<0.0001
Biosynthesis of alkaloids derived from ornithine	<0.0001	<0.0001
Lysine and nicotinic acid	<0.0001	<0.0001
Colorectal cancer	<0.0001	<0.0001
Biosynthesis of alkaloids derived from shikimate pathway	<0.0001	<0.0001
Biosynthesis of phenylpropanoids	<0.0001	<0.0001
Biosynthesis of alkaloids derived from histidine and purine	<0.0001	<0.0001
Renal cell carcinoma	<0.0001	<0.0001
Biosynthesis of terpenoids and steroids	<0.0001	0.000120652
Hematopoietic cell lineage	<0.0001	0.000156804
Biosynthesis of alkaloids derived from terpenoid and polyketide	<0.0001	0.000557965
Melanoma	<0.0001	0.000685036
Complement and coagulation cascades	<0.0001	0.000708776
Fatty acid biosynthesis	0.00014198	0.000963969
Endocytosis	0.00016526	0.001065927
Adipocytokine signaling pathway	0.00018528	0.001138149
Pathways in cancer	0.00021031	0.001191007
PPAR signaling pathway	0.00021235	0.001191007
ECM-receptor interaction	0.00026501	0.001373592
Biosynthesis of plant hormones	0.0002662	0.001373592
MAPK signaling pathway	0.00037278	0.001849562
Vitamin B6 metabolism	0.00062582	0.002853436
ErbB signaling pathway	0.00062665	0.002853436
Pyrimidine metabolism	0.00064147	0.002853436
TGF-beta signaling pathway	0.00096778	0.004161454
Vibrio cholerae infection	0.0020501	0.008531061
Toll-like receptor signaling pathway	0.0025301	0.01019947
Atrazine degradation	0.0043247	0.01690565
SNARE interactions in vesicular transport	0.0056832	0.02156273
Metabolic pathways	0.0060293	0.02222228
Tight junction	0.013217	0.04736092

Table 4.4. Significantly dysregulated pathways in our probeset.

Because many molecules are involved in several pathways, I calculated the percentage of overlap between each pathway by calculating the number of overlapping genes between each pathway and dividing it by the total number of genes. As shown in Figure 4.27, there was a significant overlap between various metabolic pathways, in addition to an overlap between

Pathways in cancer and Endocytosis, Colorectal cancer, Focal adhesion, Adherens junction and Epithelial cell signalling in H. pylori.

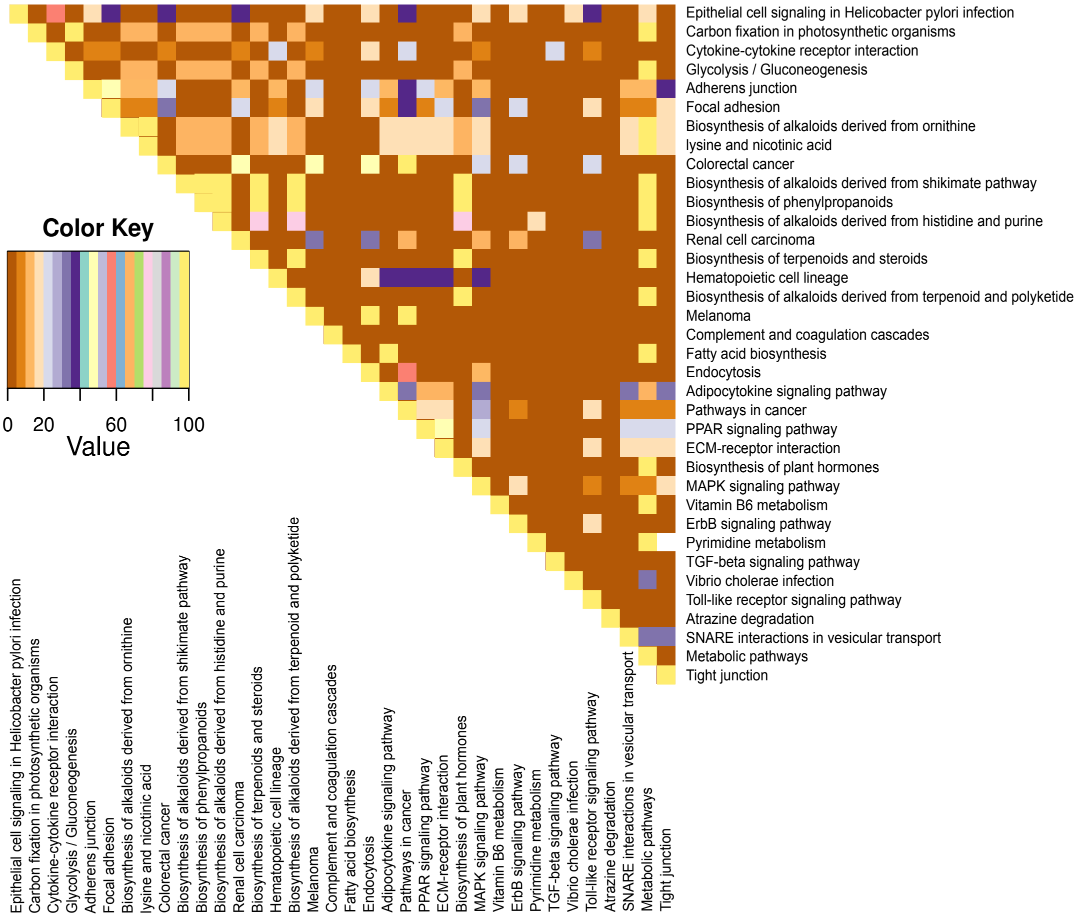


Figure 4.27 Percentage of overlap between individual pathways in genes represented in those pathways. The highest level of overlap is shown between various metabolic pathways.

4.3.6.2. TGF-beta pathway

TGF-β pathway plays a critical role and cell proliferation and differentiation, and its disruption can have both tumour-promoting and tumour-suppressive effects (Derynck et al., 2001). Additionally, mutations in TGFBR1 have been identified in multiple self-healing squamous epithelioma (Ferguson-Smith disease), a rare form of familial skin cancer that causes spontaneously regressing cSCCs (Reinhart et al., 2000). Figure 4.28 illustrates up- and down-regulated genes involved in TGF-beta signalling that have shown differential expression with adjusted p value < 0.01.

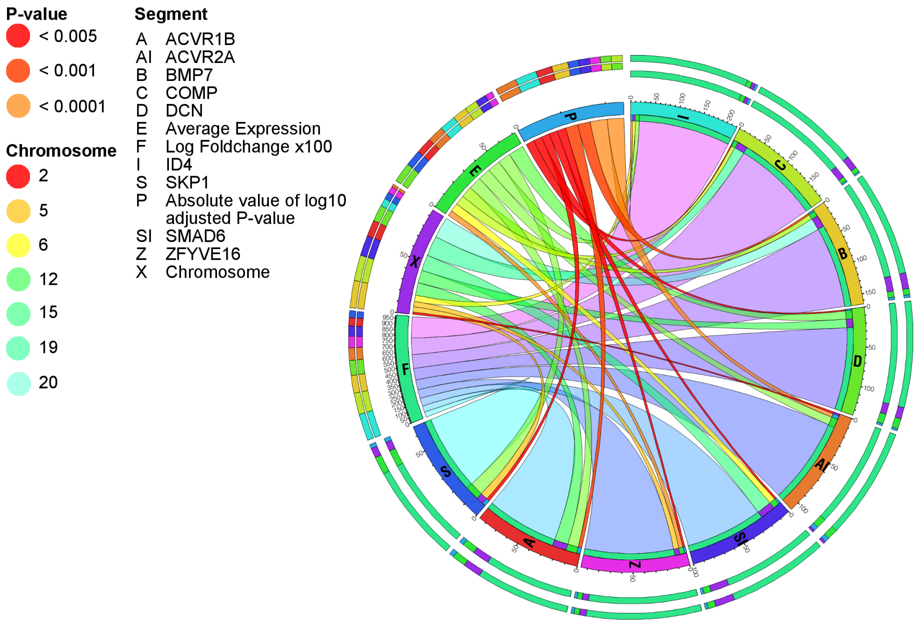


Figure 27a. Down-regulated TGF-beta pathway members.

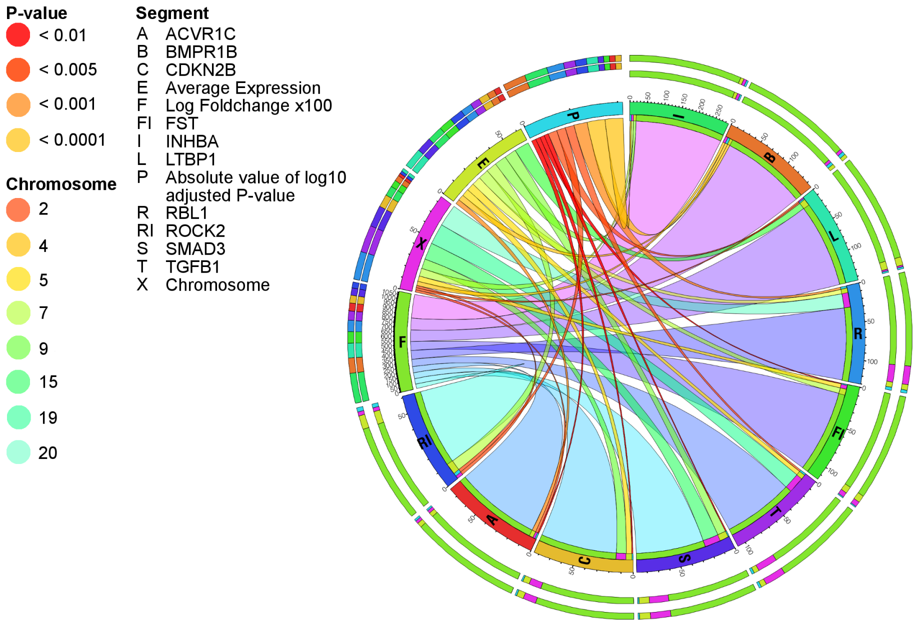


Figure 27b. Up-regulated TGF-beta pathway members.

Figure 4.28 Dysregulated TGF-beta pathway members with corresponding p value, chromosome, fold change and average expression. Up-regulated genes were generally more significantly dysregulated compared to down-regulated genes. Average expression=log2 transformed expression values, p value=log10 transformed p value (the wider the band, the more significant the p value), COMP gene=cartilage oligomeric matrix protein, DCN gene=decorin, FST gene=follistatin. X=chromosome.

A detailed look at mechanistic links in TGF-beta signalling reveals significant down-regulation of IL4, a putative tumour suppressor gene (Cabral et al., 2003), significant up-regulation of JUN oncogene (Figure 4.29).

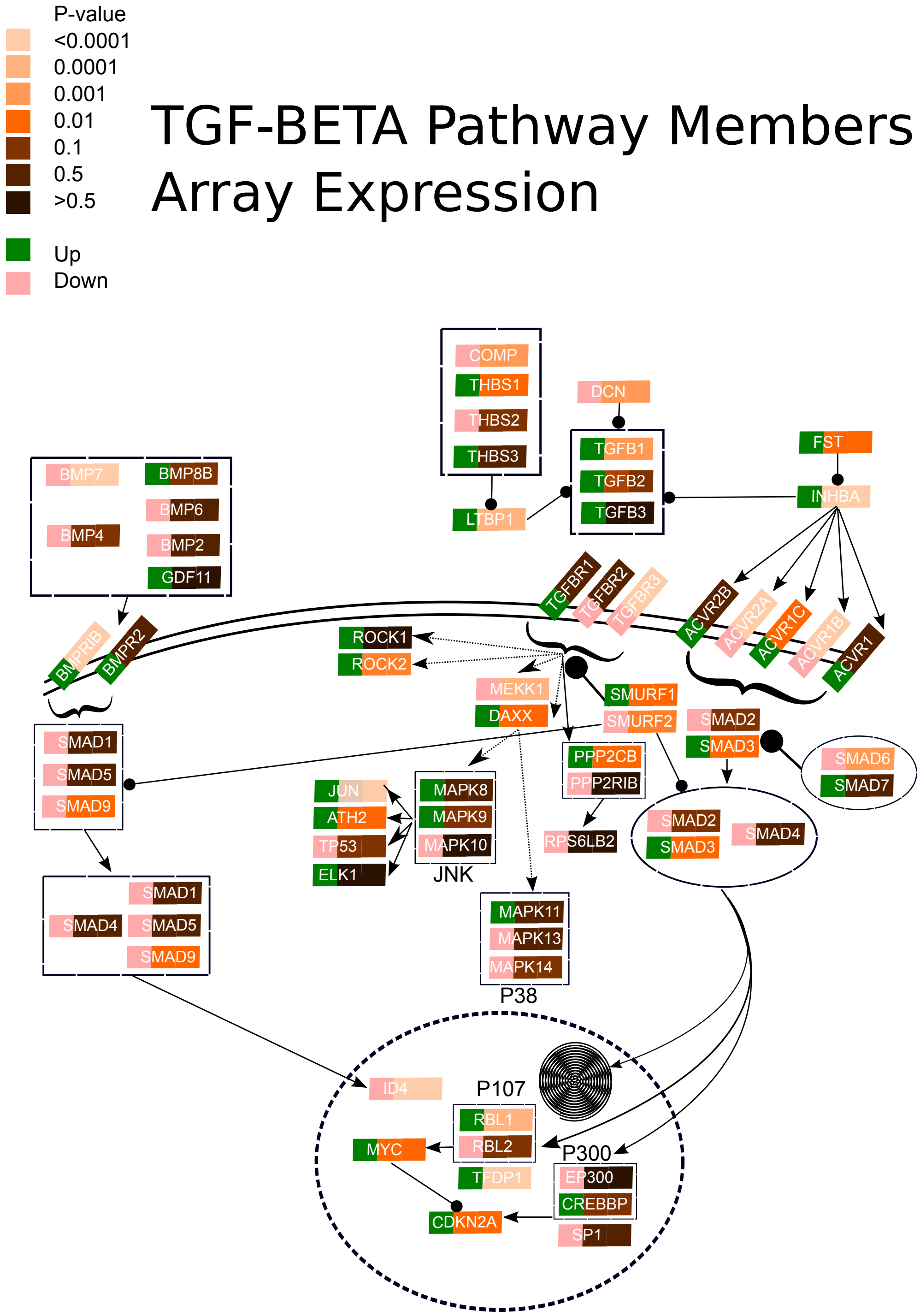


Figure 4.29 TGF-beta pathway members and their respective p values and fold change in cSCC. This figure represent KEGG-annotated TGF-beta pathway members and their interactions. Grey spiral=DNA, surrounding dashed circle=nucleus, double arched line=cellular membrane. Arrows

indicate stimulation/upreguation, black dots indicate inhibitory effect of the linked gene. This schema provides an overview of mechanistic links between TGF-beta pathway members and various critical cellular regulators including oncogens (JUN) or tumour suppressors (IL4).

4.3.6.3. RAS/MAPK pathway

This well-characterised signalling pathway plays a critical role in development (Sundaram, 2006), and its dysregulation has been linked to a variety of malignant tumours (Dhillon et al., 2007). Figure 4.30 shows upregulation of MYC (c-Myc), as downstream oncogenic effector of this pathway in our dataset, possibly due to RRAS2 signalling and PRKCA and PRKACA upregulation. Protein kinase C-α (PRKCA) is an EMT promoter (Kyuno et al., 2013), and protein kinase, cAMP-dependent, catalytic, alpha (PRKACA) has been previously shown to upregulate c-Myc through PLK1 (Padmanabhan et al., 2013). This pathway is not dysregulated in SE and NSE to AK progression (Appendix 6).

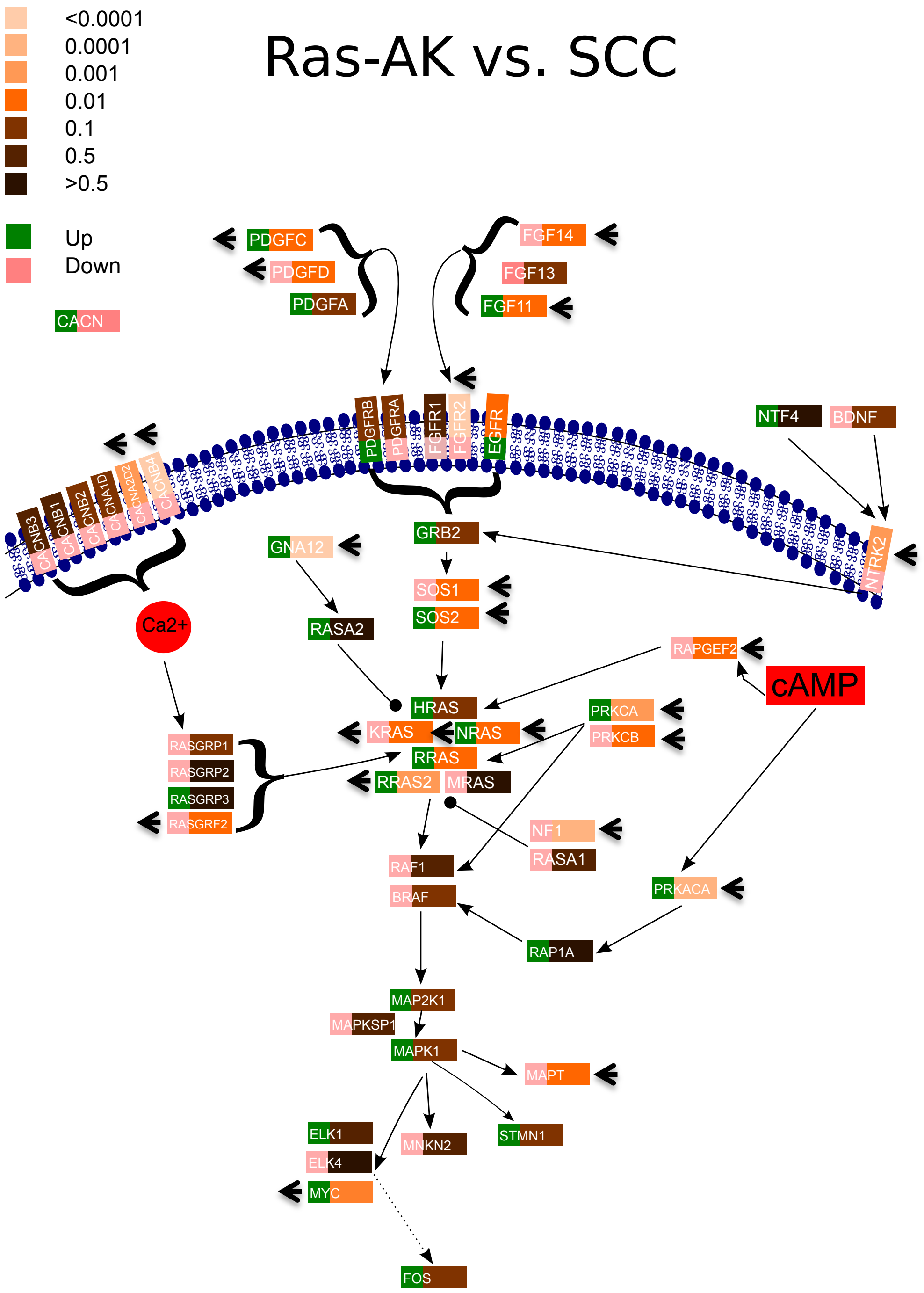


Figure 4.30 Ras/MAPK pathway signalling with corresponding p value and fold change in cSCC of relevant molecules in our dataset. This figure represent KEGG-annotated TGF-beta pathway members and their interactions. Arrows indicate stimulation/upreguation, black dots indicate inhibitory effect of the linked gene. Significantly dysregulated genes are marked with a black, arrow-like shape. This schema provides an overview of mechanistic links between the RAS pathway members and various critical cellular regulators including oncogens (MYC).

4.3.7. Integration of expression changes with previously detected genetic aberrations

Previous microarray studies of cSCC have not systematically correlated gene expression with copy number changes. To address this, regions of copy number aberration previously identified in a high resolution SNP microarray study of cSCC in our laboratory in a seriesof 60 cSCC (Purdie et al., 2009) were used to identify up- or downregulated genes located in regions of common (10% or greater) amplification or deletion, respectively. Table 2.2 provides an overview of genes that were found within a location of previously detected SCNA. Of 346 DEG, 236 (68%) could be localised within a region of SCNA. Of these, 145 (57%) genes significantly dysregulated at the expression level were found to be in corresponding regions of somatic copy number change and are highlighted in table 2.2.

Regions with the highest concordance between genetic aberration and expression change included 1pter-1p32.3 (88%), 7pter-7pcen (85%), 11q12.2-11q13.4, 13qcen-13qter, 14q11.2-14q13.1, 18q12.1-18qter (all 100%) and 19pter-19pcen (87.5%). Conversely, low concordance was observed in regions 9pter-9pcen (12.5%), 15qcen-15q15.3 (0%), 19pter-19p13.2 (20%) and 19qcen-19q13.33 (14%). It is plausible that regulation of expression within these regions may be subject to epigenetic regulation.

Location	**SCNA**	**No. of samples**	**Genes**
1pter-1p32.3	G	6	**SH2D5 CAP1 ENO1 LRRC42 SNRNP40 KTI12 SRM CLIC4 ZNF593** LOC727770 **MFAP2 FBLIM1 PLOD1 SLC9A1** SPATA6 **KIAA0090 TCEB3**
1qcen-1qter	G	8	LGR6 **LAMC2 ZNF281 ISG20L2 S100A11 NME7** LCE1B LOR SLC30A1 **C1orf74** C1orf96 FLG FLG2 **S100A10 SHC1** SERTAD4 **PPPDE1 UCK2 XPR1 PFDN2** RGS5 SLC30A1 **RAB3GAP2** DEGS1 CAPN8 LCE2B **TAGLN2**
2q21.2-2qter	L	6	**DAPL1 IGFBP5 ACVR2A SLC16A14** KLF7 ZAK DTYMK **CACNB4 FRZB TANC1** NRP2
3qcen-3qter	G	23	BOC **PDIA5** APOD AADACL2 **IGF2BP2** C3orf57 GRAMD1C **CD47 TNFSF10 RUVBL1 FNDC3B PAK2** SGEF **IL1RAP** PPM1L TMEM45A
3pter-3pcen	L	21	CDCP1 **FAM107A ROBO2** FLNB **MAGI1** RFT1 BHLHE40 HRH1 **SATB1**
4pter-4p13	L	6	HS3ST1 WDR1 **ATP8A1**
4q21.21-4qter	L	6	**HPGD MFAP3L SYNPO2 PGRMC2** MGC45800 BMPR1B **ELOVL6** SPP1
5pter-5pcen	G	12	PDZD2 ADCY2 **TARS TRIP13 CCT5 TRIO** PRLR **LPCAT1** SEPP1
5qcen-5qter	L	10	**GLRX MARCH3 C5orf46** TCOF1 **C5orf27** RNF180

			G3BP1 C5orf24
6pter-6p21.1	G	6	ID4 **SERPINB1** TMEM170B **KIAA1949** CMAH KIAA0240
7pter-7pcen	G	8	AZGP1 **INHBA GNA12 H2AFV UPP1 HEATR2 YKT6**
7qcen-7q11.22	G	6	
8pter-8pcen	L	14	**LONRF1** PROSC EIF4EBP1 CTSB ADAM9 **LZTS1** NRG1 **EGR3** BMP1
8qcen-8qter	G	11	**FAM83A PDP1** MATN2
9pter-9pcen	L	23	ANXA1 **GGTA1** LRRC8A MAPKAP1 CTSL1 PRPF4 VAV2 C9orf30
9qcen-9q21.11	L	6	**TYRP1** B4GALT1
9qcen-9qter	G	15	TYRP1 **B4GALT1**
10pter-10qter	L	7	**FAM13C** PLAU **PDCD4** KLF6 **NCRNA00081 FGFR2 LOC283070 SORBS1** NOLC1 **ACADSB**
11pter-11pcen	L	8	C11orf17 **SCUBE2BBOX1** NRIP3
11q14.1-11qter	L	6	MMP10 **SIDT2** MMP1 MMP3 **CADM1CDON SIDT2 ELMOD1**
11q12.2-11q13.4	G	6	***RPS6KA4 STIP1***
13qcen-13qter	L	11	**EDNRB ITGBL1 FREM2 UBL3 MCF2L TNFRSF19 SCEL N4BP2L1**
14q11.2-14q13.1	L	6	**STXBP6**
15qcen-15q15.3	L	6	SCG5
15q25.3-15q26.2	G	6	
16p13.11-16pcen	L	6	**GDE1** IL4R VKORC1
17pter-17qter	L	7	**GAS7** HN1 **HLF KRT9 RHOT1** SLC16A3 CSNK1D **MBTD1 ZNF652 VAMP2 KRT10** THOC4 SKA2 SPHK1 **KRT23**
18q12.1-18qter	L	6	**SERPINB1 BCL2 KIAA1632**
19pter-19p13.2	L	6	DAPK3 **ACER1** TCF3 LMNB2 SBNO2
19pter-19pcen	G	6	**GLT25D1 TPM4 DDX39 SBNO2 LMNB2 DAPK3** ACER1 **TCF3**
19qcen-19q13.33	L	6	PVR PLAUR CD3EAP **GLTSCR2** FTL EHD2 ZNF114
19q13.31-19qter	G	6	GLTSCR2 **FTL EHD2 ZNF114**
20pter-20qter	G	6	PCSK2 **SDC4 CDC25B NOP56 PTPN1 TPD52L2 RAB22A**
22q11.23-22qter	L	6	MYH9 YWHAH ZBED4 APOBEC3A **SOX10 BPIL2**
22qcen-22qter	G	11	**RANBP1 MYH9 YWHAH ZBED4 APOBEC3A** SOX10 **CDC45L** BPIL2
Xp22.33-Xqter	G	6	ITM2A PLP1 **SMS GNL3L PGK1 PCTK1** AR ATP7A

Table 4.5. Integration of genetic changes and expression profiling. Bold genes show concordant direction of expression change with copy-number change.

Chromosomal location of differentially expressed probes is shown in Figure 4.31.

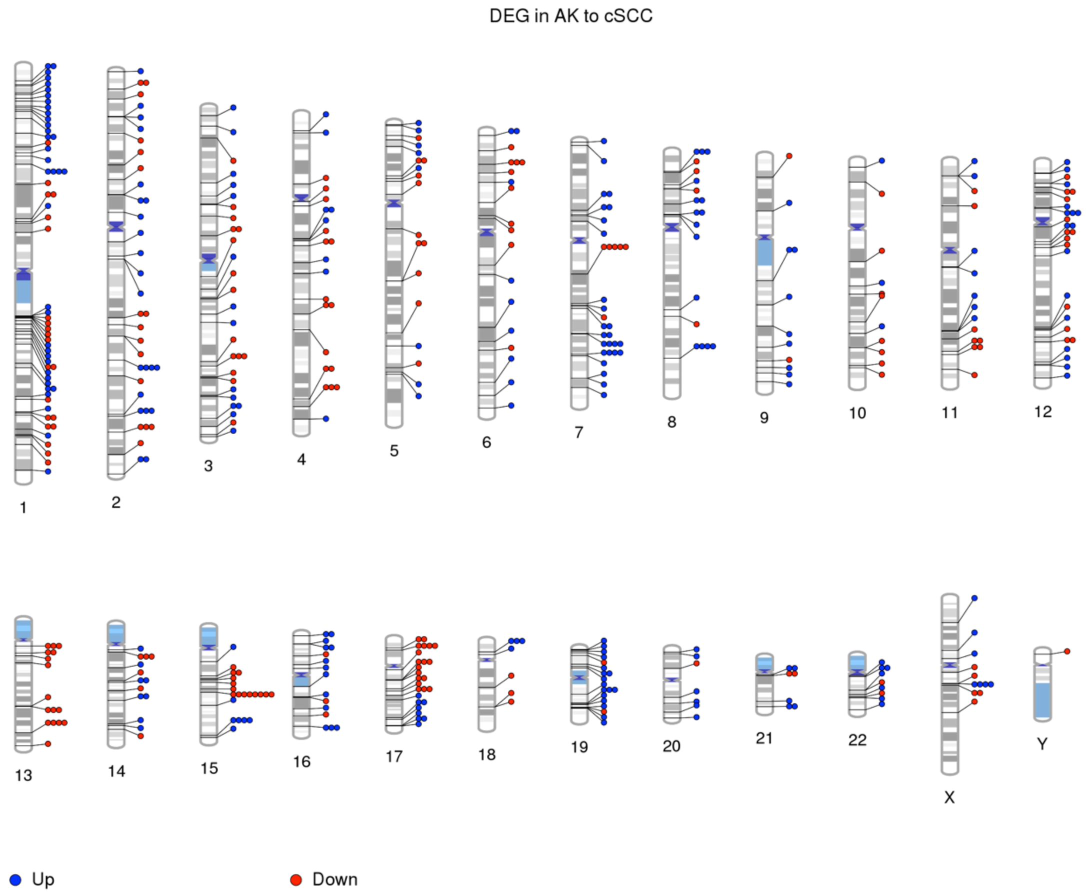


Figure 4.31 Ideogram of differentially expressed probes with known chromosomal location and respective fold change. This figure depicts regional differences in differential gene expression. For example, differentially expressed genes on chromosome 13 were all down-regulated, while genes on distant p-arm of chromosome 1 were consistently upregulated. On chromosome 18, genes located on the p-arm are up-regulated, while chromosomes on the q-arm are donwregulated.

4.3.8. Comparison with previously published datasets

4.3.8.1. Comparison with skin and non-melanoma skin cancer expression profiling data

To validate the specificity of 508 differentially expressed probes detected in our dataset, I used an independent previously published dataset (Riker et al., 2008) that consisted of 4 normal skin samples, 15 basal cell carcinoma samples, and 11 cSCC samples.

Unsupervised hierarchical clustering of this dataset post-normalisation showed clustering of 3 cSCC samples with normal skin (Figure 4.32).

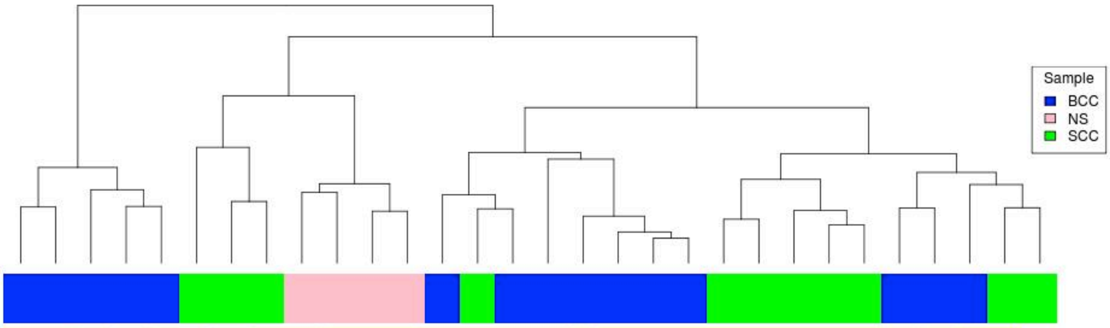


Figure 4.32 Unsupervised hierarchical clustering of normal skin (NS), cSCC and BCC samples (previously published dataset).

I then used the set of 508 probes for cluster analysis of the normalised data, which lead to a complete separation of the 11 SCC samples from 4 normal skin samples (Figure 4.33).

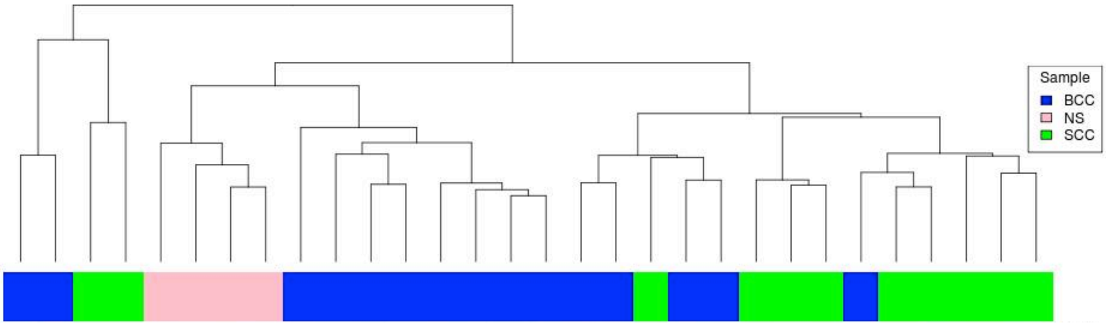


Figure 4.33 Clustering of an alternative sample set based on our top differentially expressed probes. Normal skin forms an entirely separate cluster based on these data, while in the previous figure, it clusters with 3 cSCC samples.

A comparison with another published dataset (Nindl et al., 2006) consisting of 5 cSCC samples and 3 AKs has shown that while overall transcriptional profile of these lesions is unable to distinguish between them on a cluster level (Figure 4.34), the top 508 probes derived from our dataset lead to clustering of 2/3 AK samples, further confirming the specificity of this probeset.

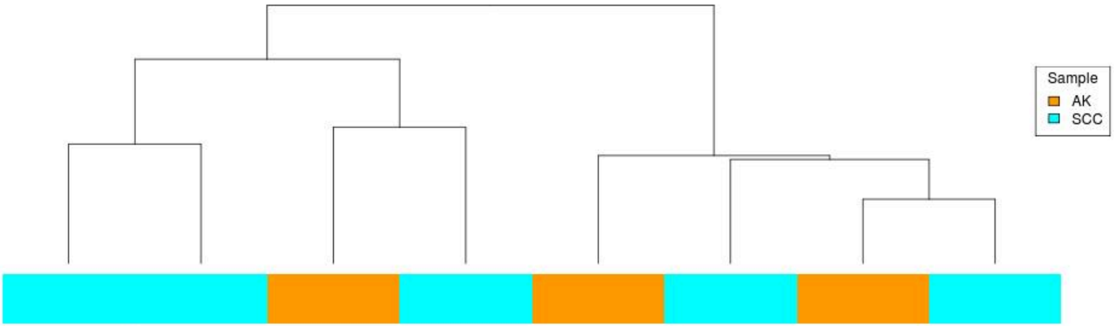


Figure 4.34 Hierarchical clustering of 8 samples from a previously published dataset. This clustering demonstrates lack of sample separation based on overall transcriptional profile.

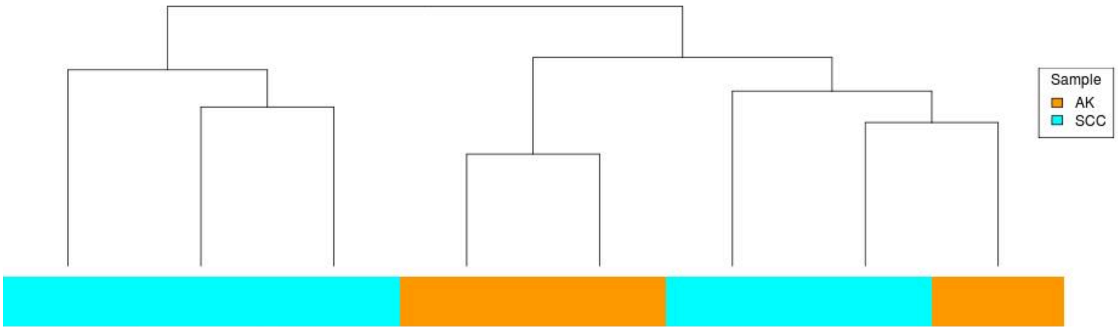


Figure 4.35 Clustering of dataset presented in the previous figure based on top 508 differentially expressed probes detected in our dataset. This probeset leads to a cluster formation of 2/3 AK samples. Given the low number of samples, it is not possible to assess statistical significance of this observation.

4.3.8.2. Comparison with stem cell data

I hypothesised that cSCC acquire partial stem-cell like transcriptional profile, and due to higher degree of dedifferentiation present in MD and PD histological subtypes, I hypothesised that such property would be more pronounced in less differentiated cSCC. I thereby compared the expression levels of 4 different mesenchymal stem cells (bone marrow, cord blood, amniotic fluid and amniotic membrane) with cSCC expression levels, and calculated the correlation between these. The highest correlation between the overall expression profile was with bone marrow stem cells (67.2%) and the lowest with amniotic membrane (66.4%). Although this difference was significant (p = 0.04, Wilcox test), ANOVA testing for all correlation values did not reach significance (p = 0.13).

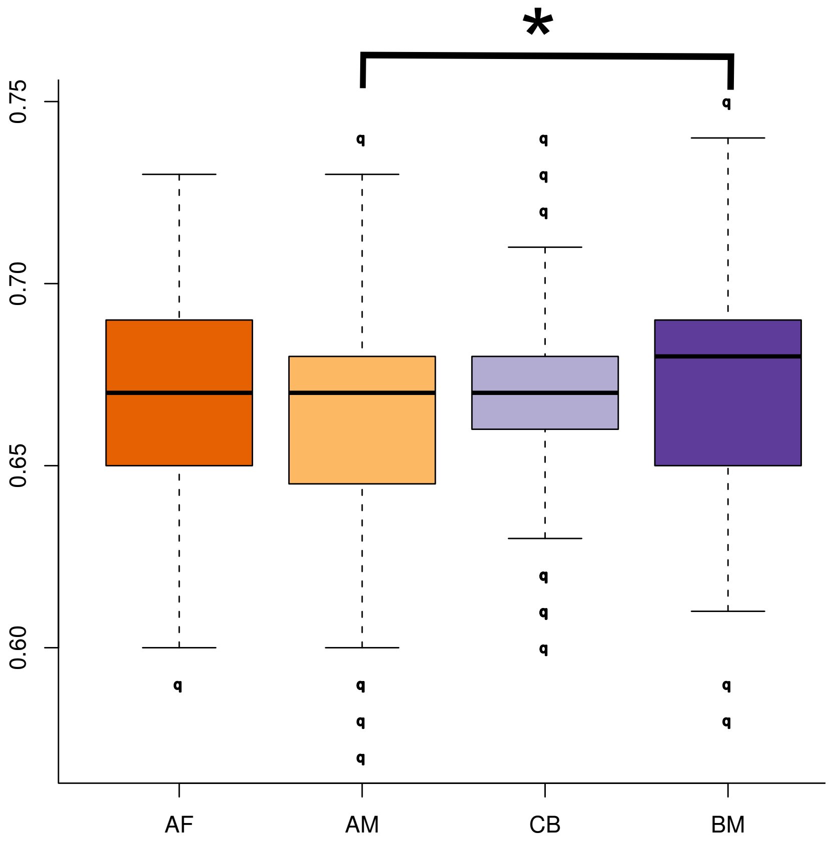


Figure 4.36 Correlation between four different MSC (amniotic fluid-AF, amniotic membrane-AM, cord blood-CB and bone marrow-BM MSC) and cSCC. BM MSC show the highest correlation. q=outliers.

Pearson’s correlation of MSC with cSCC lead to 2 clusters of cSCC (Figure 4.37). The left cluster (samples which show higher correlation with MSC) contains only 2 WD samples, while the right one contains 12 WD samples, which is a significant finding (p value 0.0111, chi square test).

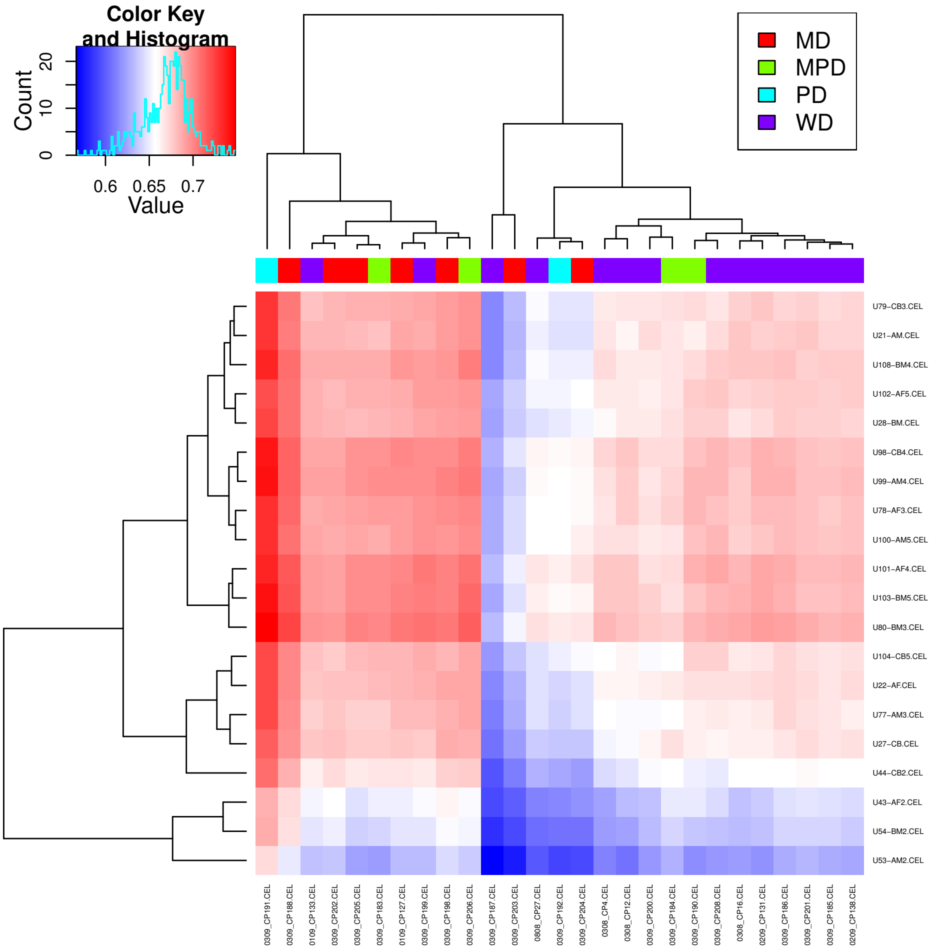


Figure 4.37 Pearson's correlation between MSC and cSCC transcriptional profile. Two clusters are formed, and the left cluster which shows higher levels of correlation contains significantly more less differentiated tumours.

4.4. Discussion and future directions

In this study, we compared the transcriptome profile of tumour cells isolated from cSCC tumours and dysplastic keratinocytes isolated from AKs, and identified differentially expressed genes (DEG) using gene expression microarrays. We validated the microarray data with an additional set of samples using QPCR.

I explored potentially dysregulated cancer pathways in the transition from AK to cSCC, and I examined the association between expression profiles and tumour histopathology, patient gender and immune status, and concordance of expression changes with known genetic aberrations. Additionally, I demonstrated that cSCC acquire stem cell-like transcriptional profile, and that transcriptional correlation with mesenchymal stem cells is significantly more pronounced in less differentiated cSCC subtypes.

4.4.1. Genome-wide transcription of cSCC and AK revealed 346 differentially expressed genes

Analysis of differential gene expression revealed 508 probes to be differentially expressed with p value < 0.0001, and this probe-set represents 346 differentially expressed genes, of which 147 (42%) are down-regulated in cSCC compared to AK, while 199 genes (58%) are up-regulated.

4.4.2. Transcriptome profiles reflect histological subtypes of cSCC

Histological typing of various tumours is not only of prognostic value, but is associated with distinct clinical characteristics and biological behaviour of cSCC. Our data show that the histopathological distinction between cSCC subtypes is not merely morphological, but is accompanied by a variation in expression profile. Taking into account previously published data from our group and others demonstrating distinct genetic profiles of high grade cSCCs (Purdie et al., 2009), it is likely that different histological grades of cSCC represent distinct disease subtypes and may require tailored therapeutic approaches based on their molecular profile. Such profiling may prove especially useful in the administration of targeted inhibitors.

4.4.3. Pathway Analysis pinpoints TGF-β and MAPK pathway as potential mediators of AK to cSCC transition

Enrichment studies of these data have detected several pathways to be dysregulated in cSCC, including pathways previously implicated in cancer. The TGF-β pathway has been previously shown to play a key role in the onset of various cancers and its pro- or anti-oncogenic role has been demonstrated to be highly context-dependent (Kukimoto et al., 2008) (Ueda et al., 2007). In addition, it has been demonstrated that this pathway promotes invasion and metastases via epithelial-mesenchymal transition (Miyazono, 2009), and has been linked to skin cancer in mice model (Eichberger et al., 2004). The BMP segment of TGF-beta pathway is consistently down-regulated in our data, while the TGF-beta-SMAD signalling part shows expression level inconsistencies, but ultimately leads to up-regulation of JUN and MYC oncogenes. ID4, a putative tumour-suppressor (Cabral et al., 2003) (Beck et al., 2007), was highly down-regulated in our dataset. ID2 down-regulation has been shown to lead to EMT induction by E-cadherin down-regulation (Yan et al., 2013), and similar mechanism may relate also to ID4, although data describing this mechanism are so far lacking.

In epidermis, TGF-beta signalling has been linked to EMT via HEY1 activation (Zavadil et al., 2004), but the expression levels of HEY1 in our dataset shows no significant difference. Although the role of TGF-beta in in multiple self-healing squamous epithelioma (Ferguson-Smith disease), has been previously characterised (Reinhart et al., 2000), the role of this pathway in sporadic cutaneous squamous cell carcinoma has not been fully clarified yet, and our data warrant further studies of TGF-beta signalling involvement in AK to cSCC transition.

The role of MAPK/Ras signalling pathway in cSCC has been previously implicated (Barnes et al., 2003); and recent clinical data indicate that a crosstalk between Ras and MAPK pathway may be responsible for cSCC development in patients treated with BRAF inhibitors (Su et al., 2012a). Our data suggest that EGFR, RRAS2, PRKCA and PRKACA upregulation may lead to increased c-Myc activity, and this regulation is strongly tumour-intrinsic and absent in normal skin to AK progression.

Erbb pathway has also been shown to be linked to EMT transition via MMP9 upregulation, and MET overexpression, a molecule which subsequently leads to EMT. MET and MMP3, both linked to EMT, were unequivocally upregulated within the top 508 probes, and MMP9 was upregulated in our data (p= 0.003). So far, there are only anecdotal data describing EMT in cSCC (Nakamura and Tokura, 2011), and since many additional cell-cell contact pathways are disrupted according to our data (Adherens Junctions, Tight Junctions), which are important in EMT, our data warrant further studies of this process in cSCC pathogenesis.

4.4.4. The role of stem cells markers in AK to cSCC transition

I detected up-regulation of several genes previously shown to be highly up-regulated in mesenchymal stem cells (PLAU, PLAUR, ANXA1, INHBA, SERPINE1, HRH1, IGF2BP2). I also showed that various markers of epidermal differentiation such as KRT9, FLG, FLG2, LOR, LCE1B, LCE2B, SERPINA12, or SCEL were significantly down-regulated in cSCC samples. These data suggest that the oncogenic process of AK transition into cSCC involves several degrees of de-differentiation, and that keratinocytes undergo at least partial epithelial-to-mesenchymal transition (EMT). Although neither TWIST1, SNAI2 nor ZEB1 were significantly differentially expressed in our dataset, all showed a trend towards up-regulation with increasing cSCC grade. Additional genes involved in EMT (claudin 11 [CLDN11] and cadherin 19, type 2 [CDH19]) shown consistent down-regulation in progression from normal skin to cSCC, which was statistically significant.

When assessing the expression level of epidermal cancer stem cells-related molecules, several genes previously shown to be upregulated in cancer stem cells were also significantly up-regulated in our dataset (MAP4, PDPN, Tenascin, RUNX1), which indicates that their upregulation may not be epidermal cancer stem cells specific. Other markers such as AXL tyroine kinase showed no significant difference between AK and cSCC (p= 0.46).

Various other stem cell markers were present on the array and their expression levels are clearly tissue-specific, since 5 AKs form a cluster separating these from cSCC, and WD tumours also form a separate cluster when expression levels of stem cells markers are compared.

When assessing the similarity between expression profile of cSCC and 4 different mesenchymal stem cells, bone marrow derived stem cells show the highest similarity. The level of similarity to mesenchymal stem cells is relatively consistent across WD tumours, and seems to be consistently higher among higher-grade tumours.

Taken together, these data suggest that malignant keratinocytes gain partial stem cells properties, which may then at least partially drive their malignant profile.

4.4.5. Integration of SCNA and gene expression data

Expression data were integrated with known regions of genomic imbalance in cSCC, and revealed regions with the highest concordance between genetic aberration and expression change, including 1pter-1p32.3 (88%), 7pter-7pcen (85%), 11q12.2-11q13.4, 13qcen-13qter, 14q11.2-14q13.1, 18q12.1-18qter (all 100%) and 19pter-19pcen (87.5%). Conversely, low concordance was observed in regions 9pter-9pcen (12.5%), 15qcen-15q15.3 (0%), 19pter-19p13.2 (20%) and 19qcen-19q13.33 (14%). Previous works integrating copy-number variation and gene expression profiling in various malignancies reported various levels of concordance between the loss or gain of genomic material and corresponding gene expression. A study of myelodysplastic syndrome, a group of premalignant conditions with high risk of malignant transformation reported a various range of altered gene expression and copy-number variation correlation, ranging from 0% to 100% (Jensen et al., 2009). Another study of multiple myeloma, reported a very high correlation between genomic LOH and gene expression changes (Lu et al., 2005), which suggests that the level of correspondence between expression and copy-number variation data is likely tumour specific, and possibly depends on selected methodology.

With respect to regions identified as demonstrating low correlation in our dataset, it is plausible that regulation of expression within these regions may be subject to epigenetic regulation, either due to changes in methylation or miRNA expression. Alternatively, this may be explicable by comparing expression data with SNP array data from only partially matching tumours.

4.4.6. Therapeutic implications

Significantly dysregulated genes identified in this study include several which are already being targeted therapeutically in other malignancies. For example, several oncogenes are up-regulated in cSCC compared to AK; inhibitors to PLK are currently being tested in clinical trials (Degenhardt and Lampkin, 2010), and MET inhibitors in pre-clinical studies (Tu et al., 2010). Additionally, ADAM10 and ADAM17 are important in maintaining epidermal homeostasis via EGFR signalling, and the over-expression of both molecules has been previously linked not only to various types of cancer, but also to various skin conditions including atopic dermatitis (ADAM10), and psoriasis (ADAM17) (Tanemura et al., 2005). This observation likely links both molecules with more general processes such as inflammation and keratinocyte proliferation, rather than specifically with malignant transformation. Both molecules show a high degree of similarity in their catalytic site structure with other metalloproteinases (MMPs). Although clinical trials of metalloproteinase inhibitors have proved unsuccesful to date, further selective inhibitors are being developed (Saftig and Reiss, 2011). Once those molecules are available, cSCC represents a relevant candidate for preclinical testing based on our results. In addition, many pro-metastatic molecules were up-regulated in our dataset, such as PTHLH / osteopontin (SPP1), representing attractive therapeutic targets in advanced disease and in prevention of metastatic disease spread. Simultaneous inhibition of VEGF and osteopontin has already been proved beneficial in hepatocellular carcinoma during *in vitro* testing (Kou et al., 2010).

Aurora kinases are serine/treonine protein mitotic kinases that play an improtant role in different stages in cell division. So far, aurora kinase family inclues Aurora A, B and C, and all members are highly conserved across species. RCC2 (Regulator of chromosome condensation 2) is a passenger protein that is involved in alignment of chromosomes on the spindle and centromeric targeting of the chromosome passenger complex (Humphries et al., 2009). It is also important in Aurora B activation (Rosasco-Nitcher et al., 2008).

In our expression data, we found a significant upregulation of RCC2 (4.5 fold upregulation, pvalue= 0.0004), also, upregulation of Aurora B (3.25 fold upregulation, p value =0.01). Additionally, Aurora B is regulated by INCENP that was also upregulated (4.26 fold upregulation, p value = 0.0009). Aurora A was also upregualted in our dataset (4.1 fold upregulation, p value = 0.001). Those data suggest that Aurora kinases are significantly upregulated in cSCC compared to AK, and given that there are now available targetted therapies for Aurora kinases that are being clinically tested, these represent ideal targets for futher exploration in cSCC with strong translational potential.

4.4.7. Strengths and limitations of our study

The main advantage of our study is that the transcriptional profile is derived almost exclusively from malignant cells and dysplastic keratinocytes isolated from tumours and AKs using meticulous laser-capture microdissection. Thus, potentially misleading expression levels from stromal leukocytes or non-tumour blood vessel cells should be minimised in our dataset. Although our QPCR validation was conducted on non-microdissected samples, it has consistently confirmed the expression pattern of our microarray data, yet the levels of transcription differences were not as striking as on the array, which may be attributable to the admixture of non-malignant cells in the validation samples.

Limitations of the study include relatively low number of AKs and also the fact that these AKs were surgically removed before their actual progression into cSCC, thus it is uncertain whether their molecular features would enable progression into malignant entities. Additionally, paired comparison of 8 AKs and 8 cSCC has shown much more limited numbers of DEG, which is possibly due to the low number of samples, or due to more even balance of the two samples. Also, the relatively low number of higher grade cSCCs (MD, MPD and PD) has made it difficult to detect specific genes potentially responsible for unique molecular properties of individual cSCC subtypes.

4.4.8. Summary and conclusions

Detailed understanding of molecular changes of AK transition to malignant cSCC has been lacking despite previous attempts to elucidate this process. In this study, we show that a complex series of changes occur at the molecular level during this progression, and involve dysregulation of several hundreds of genes. Specifically, dysregulation of several oncogenic, metabolic and regulatory pathways alongside disrupted expression of several tumour suppressors and over-expression of many oncogenes is involved in this progression. This study also shows the overall transcription profile of cSCCs differs based on tumour histology, irrespective of the gender or immune profile.

In conclusion, our data highlight several known oncogenic targets, as well as potential novel factors involved in this process (epidermal stem cell markers, EMT genes). It has uncovered a broad spectrum of disrupted molecular pathways and processes that are involved in this transition. In particular, it suggests that malignant cells undergo de-differentiation, acquiring an expression profile that leads to differences in expression of stem cells markers between precancerous and tumour lesions. Finally, our study confirms at the transcriptome level our previous data showing distinct genotype of various histopathological subtypes of cSCC.

4.4.9. Implications for future research

This study has provided unprecedented insight into trascriptional events in cSCC and has pointed towards potential regions of epigenetic regulation in the cSCC cancer genome, therapeutic targets and avenues for miRNA profiling. Exploration of these areas is discussed in the following chapters.

Additionally, the correlation of less differentiated cSCC samples with mesenchymal stem cells warrants validation with a larger set of MD, MPD and PD tumours and an additional set of mesenchymal stem cells. Identification of a “stem cell signature” may provide insight into the molecular background of shared cellular properties of MSC and cSCC tumours.

Our data also indicate that EMT plays an important role in the progression from skin to AK and then in less differentiated tumours. Conditional knock-down of the SNAI2 gene identified in our data as relevant in this process in both normal human keratinocytes and cSCC cell lines paired with ChIP-seq and transcriptional profiling would elucidate its role in the malignant transformation in human keratinocytes at molecular levels.

Further evaluation of cSCC transcriptome in comparison with fresh frozen normal skin would provide novel insight into mRNA dysregulation between clinically normal tissue and this malignancy.

5. Genome-wide methylation and transcription profiling of Cutaneous Squamous Cell Carcinoma Cell Lines

5.1. Introduction

As described in Chapter 1.3.2.2, squamous cell carcinoma of the skin is the second most common malignant tumour to affect humans preceded only by basal cell carcinoma. Its incidence has been increasing at an alarming rate during the past decades, and the neoplasm represents an imminent global health-problem associated with considerable morbidity (O'Bryan et al., 2013), health-care costs (Vallejo-Torres et al., 2013) and suffering (Koster and Bergsma, 1990). Unlike in basal cell carcinoma, in which the PTCH gene has been shown to be critical for its development, no single gene or molecular pathway has been implicated as crucial in the development of cutaneous squamous cell carcinoma (cSCC).

Epidermolysis bullosa (EB) is a group of rare inherited blistering skin diseases associated with serious body mutilations and mortality. Afflicted patients are at increased risk of developing cSCC, with metastatic cSCC representing the most common cause of death in patients reaching adulthood. Moreover, EB-associated cSCC exhibits a more aggressive clinical behaviour and is more difficult to identify.

Recessive dystrophic epidermolysis bullosa (RDEB) is caused by mutations in collagen, type VII, alpha 1 protein-coding gene. This protein is a predominant constituent of fibrils that anchor epidermal basement membrane to the underlying dermis. In RDEB, the absence or reduction of these anchoring fibrils leads to skin blistering. RDEB cSCC are at increased risk of developing aggressive cSCC (Fine et al., 2009), and tumours appears usually in skin wounds and areas affected by long-term scarring, rather than in sun-exposed areas, which indicates that different pathophysiological mechanisms play role in the onset of RDEB cSCC compared to UV-induced cSCC. Moreover, RDEB cSCC are often multiple, and 80% of patients die from cSCC metastases within 5 years following initial tumour presentation (Fine et al., 2009).

The acquisition of stem cell-like properties and dedifferentiation have recently been shown to represent the initial step in colon cancer carcinogenesis (Sauermann et al., 2008), and conversely, mesenchymal stem cells (MSC) have been shown to gain tumourigenic properties (Rubio et al., 2008), suggesting that certain molecular traits shared between the cell types may at least partially underpin oncogenic transformation.

Cancer cell lines represent a suitable system for basic and translational research (Paz et al., 2003), and previous studies focusing on aberrant methylation in cancer cell lines have identified important mechanism that underpin neoplastic growth (Herman et al., 1995). To date, no comprehensive DNA methylation or transcription study of cSCC cell lines has been conducted.

As discussed in Chapter 1.6.1, human embryonic stem cells (hESC) represent a unique type of cells capable of differentiating into any kind of body tissue. As these cells are derived from a blastocyst, all cells in the mass possess identical DNA, yet can differentiate into distinct cell types via a myriad of epigenetic and other regulatory changes. In addition to capacity for self-renewal, hESC possess migratory abilities that allow the cells to fulfil their developmental function. Such features are often ascribed to cancer cells, and various conserved signalling pathways including MAPK/ERK, PI3K/Akt, Wnt, TGF-β and NOTCH have been shown to play a critical role in both malignant tumours and hESC (Dreesen and Brivanlou, 2007).

The aims of this study included investigation of potential differences in genome-wide CpG island methylation and gene transcription between primary cutaneous squamous cell carcinoma-derived cell lines and primary cultured keratinocytes. In the view of distinct clinical characteristics of recessive dystrophic epidermolysis bullosa (RDEB)-associated cSCC, I aimed at exploring potential differences in methylation patterns between cell lines derived from cSCCs of non-RDEB and RDEB patients. Finally, I hypothesised that cSCC cells gain certain unique molecular properties normally confined to stem cells in the process of dedifferentiation that at least partially determine their malignant properties due to acquiring methylation levels of selected genes corresponding to their methylation status in embryonic stem cells.

5.2. Materials and Methods

5.2.1. cSCC cell lines and normal human keratinocytes

A total of nine cSCC lines were used for the detection of differentially methylated and transcribed genes. Three early-passage primary normal human keratinocytes derived from three different participants served as a normal control in each comparison. Additionally, 3 human embryonic stem cells (hESC) samples were used for transcriptional profiling.

Histological subtyping of original tumour with corresponding cell line, passage number and patient characterstic is presented in Table 5.1.

Illumina 270Infinium HumanMethylation27 BeadChip

Cell Line	Histological Typing	Passage No.	Patient Characteristic
NHK1	Primary normal human keratinocytes	2	Healthy adult
NHK2	Primary normal human keratinocytes	2	Healthy adult
NHK3	Primary normal human keratinocytes	2	Healthy adult
T1	Well-differentiated	5	Organ-transplant recipient
T2	Well-differentiated	13	Organ-transplant recipient
T8 3-1b	Poorly-differentiated	8	Organ-transplant recipient
IC1	Moderately-differentiated	8	Immunocompetent patient
MET1	Poorly-differentiated	11	Organ-transplant recipient
MET2	Poorly-differentiated (recurrence of MET1)	11	Organ-transplant recipient

EB2	Poorly-differentiated	13	Epidermolysis bullosa patient
EB3	Well-differentiated	17	Epidermolysis bullosa patient
GP	unknown	22	Epidermolysis bullosa patient

Table 5.1 Characteristics of cell lines hybridised to Illumina 27K methylation array.

5.2.2. DNA Extraction and Bisulphite Modification

DNA extraction from a subset of samples was conducted in collaboration with Angela McHugh[13]: DNA was extracted from confluent cell cultures at passage numbers indicated in Table 5.1 using DNesy Mini extraction kit (Qiagen, United Kingdom) according to the manufacturer's instructions. 500-1000 ng of DNA yield from each sample were bisulphite modified using the EZ DNA Methylation Kit (Zymo Research, CA, USA) according to the manufacturer's instructions in collaboration with Cambridge Genomic Services[14] and Nadiya Mahmoud[15].

5.2.3. Infinium HumanMethylation27 BeadChip

Methylation detection was performed using the Illumina Infinium HumanMethylation27 BeadChip in collaboration with Cambridge Genomic Services. This platform detects the methylation status of 27,578 CpG sites spread across 14,495 genes by sequencing-based genotyping of bisulfite-converted DNA. Bisulfite-converted (unmethylated) and unconverted (methylated) sites are evaluated by DNA hybridisation to specific probes attached to beads. There are two probes for each CpG site per locus, one for methylated and one for unmethylated sites. This is followed by allele-specific base extension with a fluorescent label, which is specific for either the methylated or unmethylated allele. Eventual methylation scores (denoted “beta-value”) are generated for each site with BeadStudio software (Illumina, Inc., USA) and range from 0 (fully unmethylated) to 1 (fully methylated). The Illumina Infinium methylation assay was performed according to the manufacturer's instructions. Briefly, bisulphite-converted DNA was amplified, fragmented and hybridised to the chip arrays, followed by imaging with the Illumina BeadArray reader. Image processing and intensity data extraction were performed according to Illumina's instructions.

5.2.3.1. Quality Control and Normalisation

Background-corrected beta-values were used to evaluate the quality of individual arrays. This was conducted using boxplots of total intensity (calculated as addition of unmethylated and methylated

---

[13] Division of Cancer Research, Medical Research Institute, Jacqui Wood Cancer Centre, University of Dundee, Dundee, Scotland.

[14] Cambridge Genomic Services, Department of Pathology, Cambridge University, Cambridge.

[15] Genome Centre, Barts and the London School of Medicine and Dentistry, London.

values [I=U+M] for each probe) shown in Figure 5.1, and histograms of beta-values on each chip (shown in Figure 5.2). If adequate bimodal distribution of beta-values was observed on histogram, the sample was included in further analysis. All twelve samples passed the quality control and were included in down-stream analysis.

Intra-array normalisation was performed in the Illumina BeadStudio software. Beta-values were further normalised using a quantile-normalisation strategy to reduce inter-array variation using preprocessCore package in R (http://bmbolstad.com/stuff/qnorm.pdf).

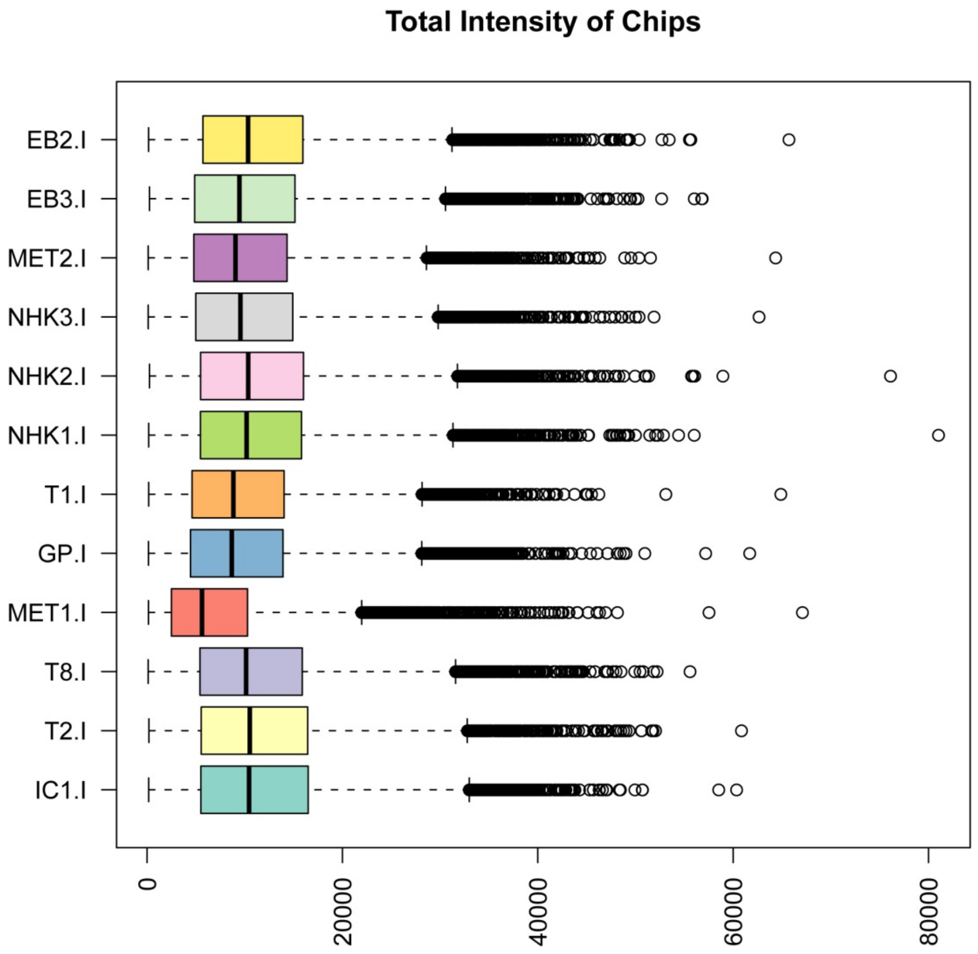


Figure 5.1 Boxplots showing total intensities of samples hybridised to Illumina 27K beadchip. X axis=total locus signal intensity (the sum of unmethylated and methylated probe intensities). MET1 sample shows generally lower intensity values compared to the rest of the samples, but was not excluded based on this quality control plot.

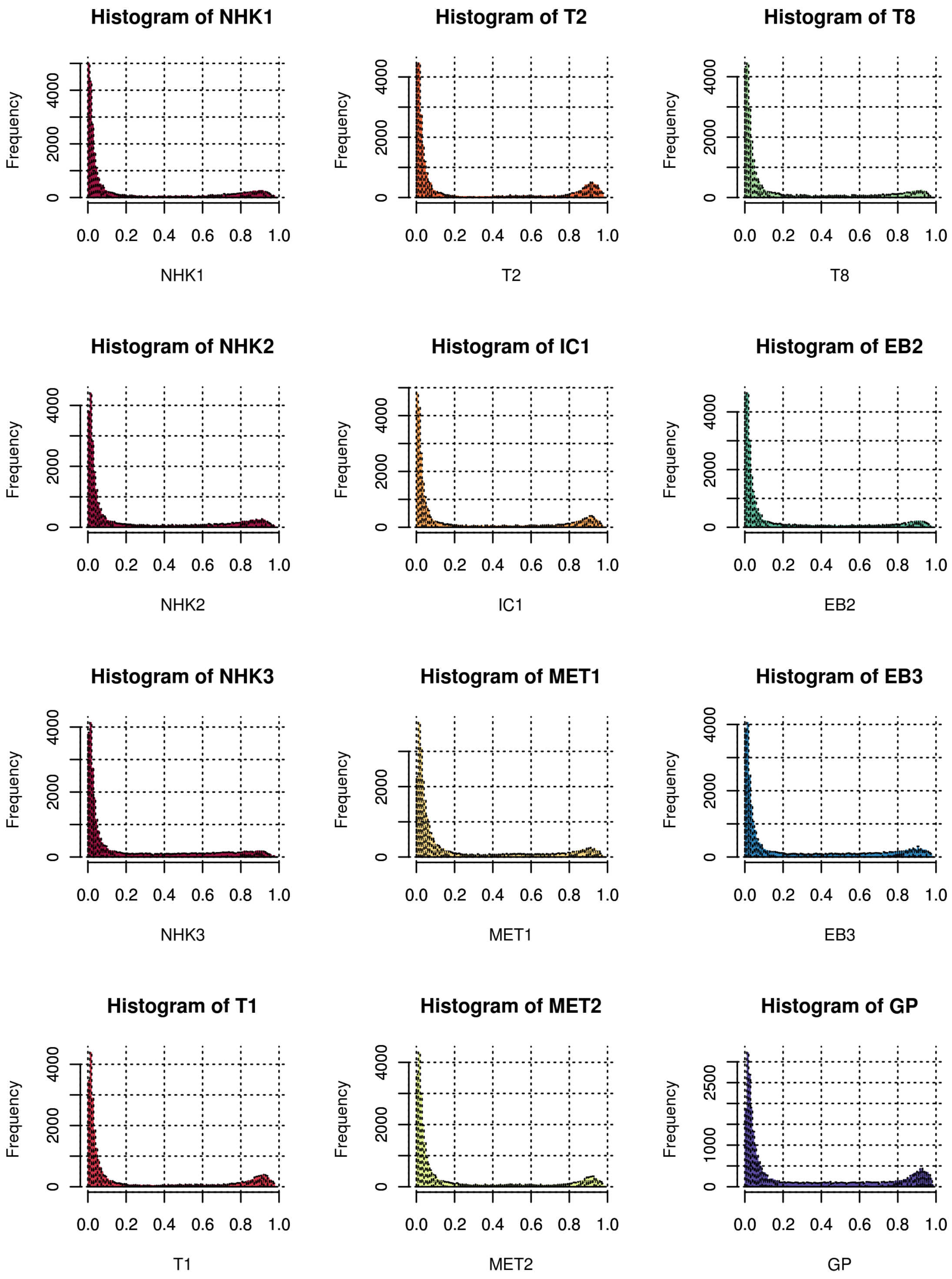


Figure 5.2 Histograms of raw, background-corrected beta-values in all samples showing appropriate bimodal distribution. This pattern confirms suitability of the data for normalisation and further downstream analysis.

5.2.3.2. Detection of Differentially Methylated Genes

Significant differences in methylation of specific loci were detected using a *genefilter* package in R. Welch two-sample test was applied to each probe to detect the difference in methylation level between the controls and all nine tumour samples. Analogous strategy was applied in detection of methylation differences between EB-derived tumours and non-EB tumours. α-level was set at pvalue<0.01. Obtained p-values were adjusted for multiple testing using Bonferroni-Holm method.

5.2.4. Comparison with previously published data

Previously published methylation profile of 30 normal skin samples obtained from healthy volunteers (Gronniger et al., 2010) using the same methylation microarray was used to validate primary cultured human keratinocytes as a suitable model for exploring methylation patterns in human skin.

Additionally, methylation profile of a panel of various cultured cell types (fetal lung fibroblasts [1x], mesenchymal stem cells [4x], cord blood CD34+ cells [2x], peripheral blood CD34+ cells [3x], bone marrow CD34+ cells [3x], peripheral blood mononuclear cells [2x], cord blood mononuclear cells [2x], human umbilical vein endothelial cells [1x], cord blood endothelial progenitor cells [1x], cord blood CD34- cells [1x], induced pluripotent stem cells [23x], embryonic stem cells [11x] and human embryonal carcinoma cells [1x], (Chou et al., 2011)) was used to determine potential stem cell-like characteristics of cSCC cell lines.

5.2.4.1. hESC signature in cSCC cell lines

I defined the epigenetic hESC signature as genes that are differentially methylated in cSCC compared to PHK that show a difference of 20% or more in averaged beta-value in hESC compared with PHK, and less than 5% methylation difference in cSCC compared with hESC. In order to identify hESC epigenetic signature in cSCC cell lines, the raw methylation data derived from 11 hESC samples were normalised with PHK and cSCC data. Probes with missing data in any of the 23 samples were filtered out. Mean methylation levels for each probe were calculated for each tissue type (hESC, PHK, cSCC). The average methylation difference in percentage points for each probe were then calculated by taking the absolute value after subtracting the average methylation levels in hESC from PHK and from cSCC (hESC minus PHK, hESC minus cSCC) and multiplying it by 100.

5.2.4.2. Average methylation differences in tissue types

The average difference in methylation across various tissue types included in a previously described dataset (Chou et al., 2011) was calculated by substracting the average methylation level for each probe between each tissue type in all permutations. Then the difference was averaged for each of those comparisons and multiplied by 100.

5.2.5. Transcriptional profiling

Illumina HumanHT-12 v4 expression array was used for this experiment to allow integration with previously used Illumina methylation array. Sample processing and RNA hybridisation was carried out in collaboration with Cambridge Genomic Services. A total of 500 ng of mRNA (RIN value >8) was amplified with the TotalPrep 96-RNA amplification kit (Applied Biosystems, USA). Briefly, mRNA is reverse-transcribed into cDNA prior to amplification, and subsequently amplified by In Vitro Transcription (IVT). This process generates biotin-labelled cRNA, which is then hybridised to the beadchip following the Direct Hybridisation assay. The microarray is read with Illumina BeadStation scanner using iScan software. Table 5.2 lists samples hybridised to the transcription profiling array.

Illumina HumanHT12v4 BeadChip 1

Cell Line	Histological Typing	Passage No.	Patient Characteristic
PK0	Normal human keratinocytes	0	Healthy adult
PK1	Normal human keratinocytes	1	Healthy adult
PKA	Normal human keratinocytes	3	Healthy adult
T1	Well-differentiated	8	Organ-transplant recipient
T2	Well-differentiated	4	Organ-transplant recipient
T8 3-1b	Poorly-differentiated	8	Organ-transplant recipient
IC1	Moderately-differentiated	8	Immunocompetent patient
MET1	Poorly-differentiated	12	Organ-transplant recipient
MET2	Poorly-differentiated (recurrence of MET1)	12	Organ-transplant recipient
EB2	Poorly-differentiated	NA	Epidermolysis bullosa patient
EB3	Well-differentiated	NA	Epidermolysis bullosa patient
GP	unknown	NA	Epidermolysis bullosa patient

Illumina HumanHT12v4 BeadChip 2

Cell Line	Histological Typing	Passage No.	Patient Characteristic
P0	Primary normal human keratinocytes	0	Male infant
PKS	Primary normal human keratinocytes	1	Healthy adult
PK5	Primary normal human keratinocytes	5	Healthy adult
T1	Well-differentiated	13	Organ-transplant recipient
T2	Well-differentiated	12	Organ-transplant recipient
T8 3-1b	Poorly-differentiated	9	Organ-transplant recipient
IC1	Moderately-differentiated	8	Immunocompetent patient
MET1	Poorly-differentiated	14	Organ-transplant recipient
MET2	Poorly-differentiated (recurrence of MET1)	13	Organ-transplant recipient

HESC1	Human embryonic stem cells	NA	Human embryonic stem cells
HESC2	Human embryonic stem cells	NA	Human embryonic stem cells
HESC3	Human embryonic stem cells	NA	Human embryonic stem cells

Table 5.2 Characteristics of cell lines and cultured cells hybridised to gene expression array.

5.3. Results

5.3.1. DNA Methylation Patterns

All twelve samples (three PHK, nine cSCC cell line samples) were hybridised to Illumina Infinium HumanMethylation27 BeadChip arrays after passing quality control and inter-array normalisation procedure to remove variation due to experimental and other confounding factors. This resulted in a normalised data matrix of beta-values between 0 and 1. Inherent bi-modality of the methylation value distribution is captured in Figure 5.3 with a typical high peak of hypomethylated loci and a low peak in the hypermethylated loci.

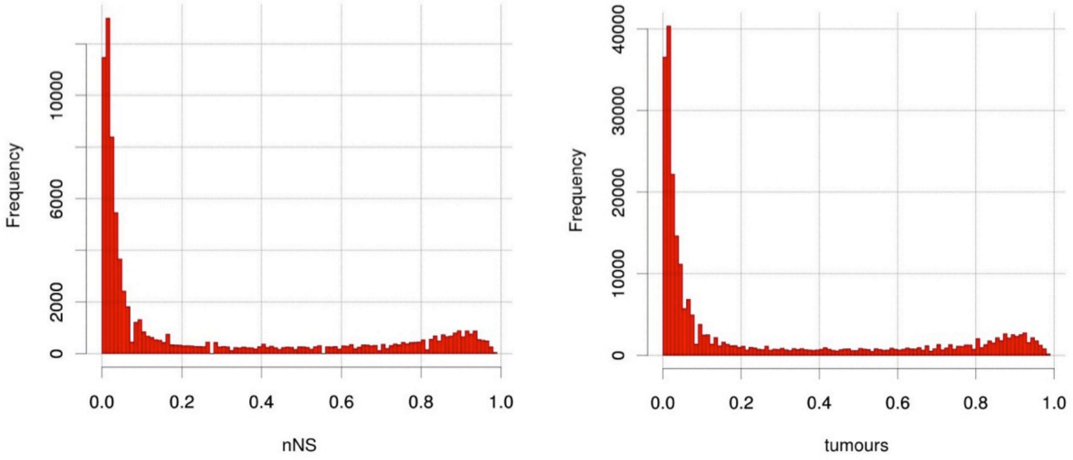


Figure 5.3 Bimodal distribution of methylation values across both sample types. This observed pattern is expected given the character of the methylation array. nNS=primary human keratinocytes, tumours=cSCC cell lines.

5.3.2. Clustering Analysis of normalised methylation profiles

Methylation profiles of all samples were correlated using Pearson's correlation algorithm, and unsupervised hierarchical clustering of these data lead to a clear separation of RDEB-derived tumours from the remaining samples (Figure 5.4). Additionally, normal human keratinocytes also formed a cluster, and non-RDEB cSCC cell lines clustered based on the original tumour histology.

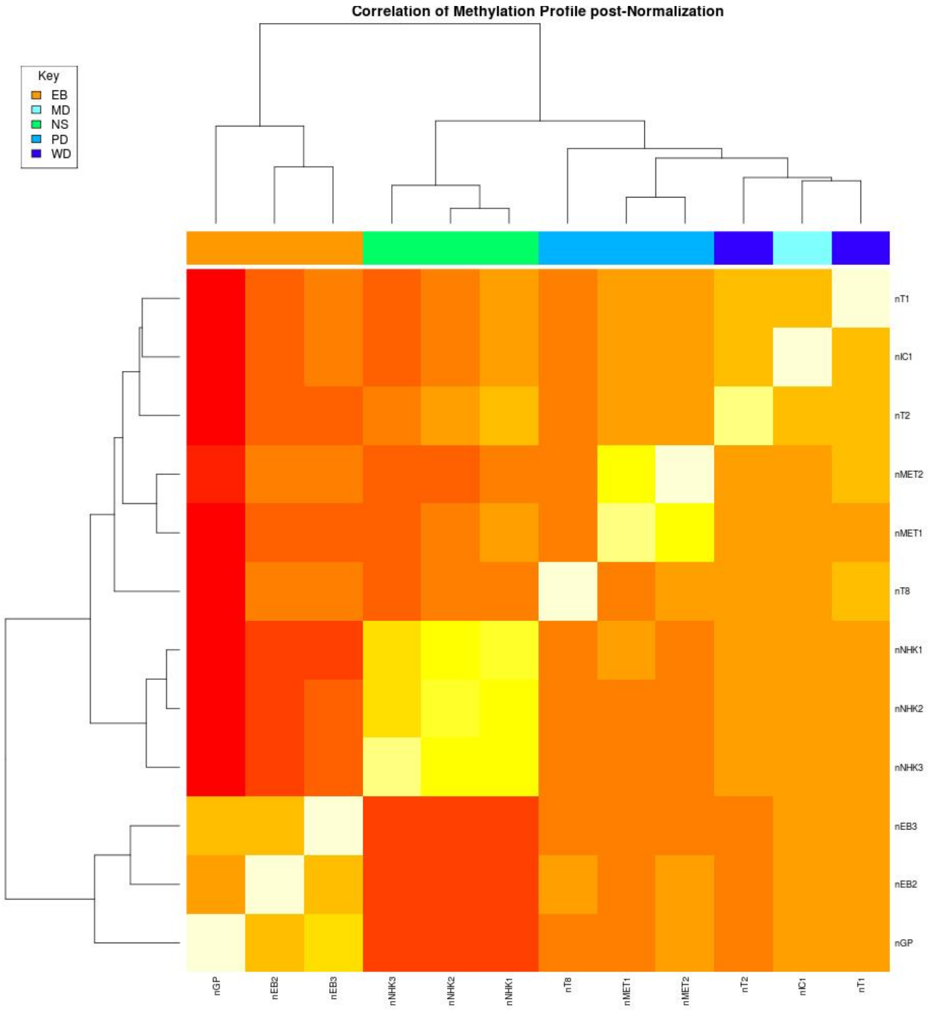


Figure 5.4 Unsupervised hierarchical clustering analysis of normal human keratinocyte samples and cSCC cell lines based on Pearon's correlation of their methylome. This graph shows clear separation based on the patient source of the original tumour (RDEB, non-RDEB patients), and on the original histology. EB=RDEB-cSCC derived cSCC cell line, MD=moderately differentiated cSCC, WD=well-differentiated cSCC, PD=poorly differentiated cSCC, NS=primary human keratinocytes.

To further explore these clusters and to quantify the statistical relevance of these, I plotted a bootstrap correlation plot shown in Figure 5.5. This figure demonstrates that RDEB-derived samples and primary human keratinocytes are entirely distinct from non-RDEB cSCC cell lines. Among these cell lines, MET1 and MET2 cell lines form an unequivocal cluster, as do cell lines derived from 2 WD and 1 MD samples (T1, T2, IC1).

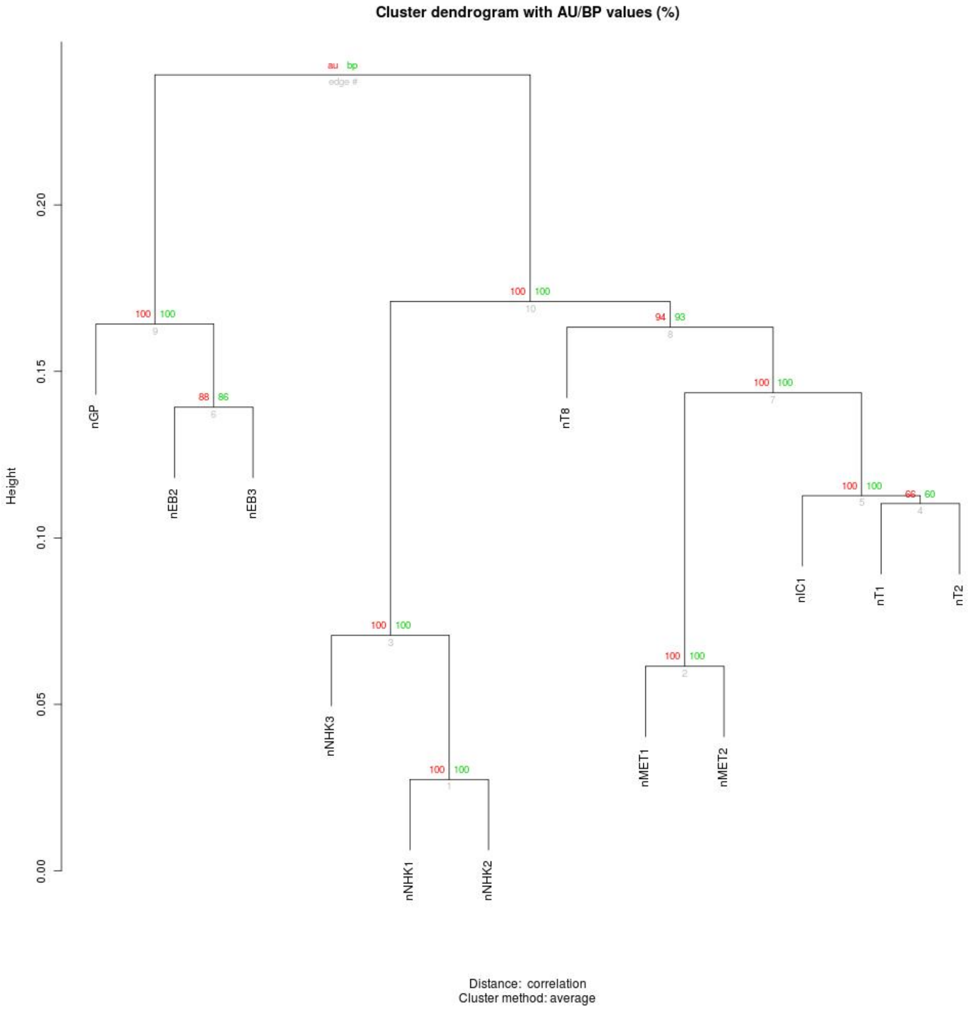


Figure 5.5 Bootstrap correlation dendrogram quantifying the statistical significance of observed clusters. Red value >95 indicates a statistically significant cluster. This figure shows that all RDEB-cSCC cell lines, NHK and 5/6 cSCC cell lines form statistically significant clusters based on their methylome.

5.3.3. Differentially Methylated Genes

5.3.3.1. Methylation levels in primary human keratinocytes compared with cSCC cell lines

With set α-level, the comparison of normal keratinocytes with cSCC cell lines identified 398 differentially methylated probes (full list is provided in Appendix 7). A heatmap of this probeset's beta-values is depicted in Figure 5.6, and shows a distinct clustering pattern of the 2 sample types based on this probeset. A histogram of adjusted p-values for all probes is shown in Figure 5.7 (top panel), indicating that the vast majority of probes present on the array are not differentially methylated. P-values for each probe with corresponding chromosomal location are plotted in Figure 5.8.

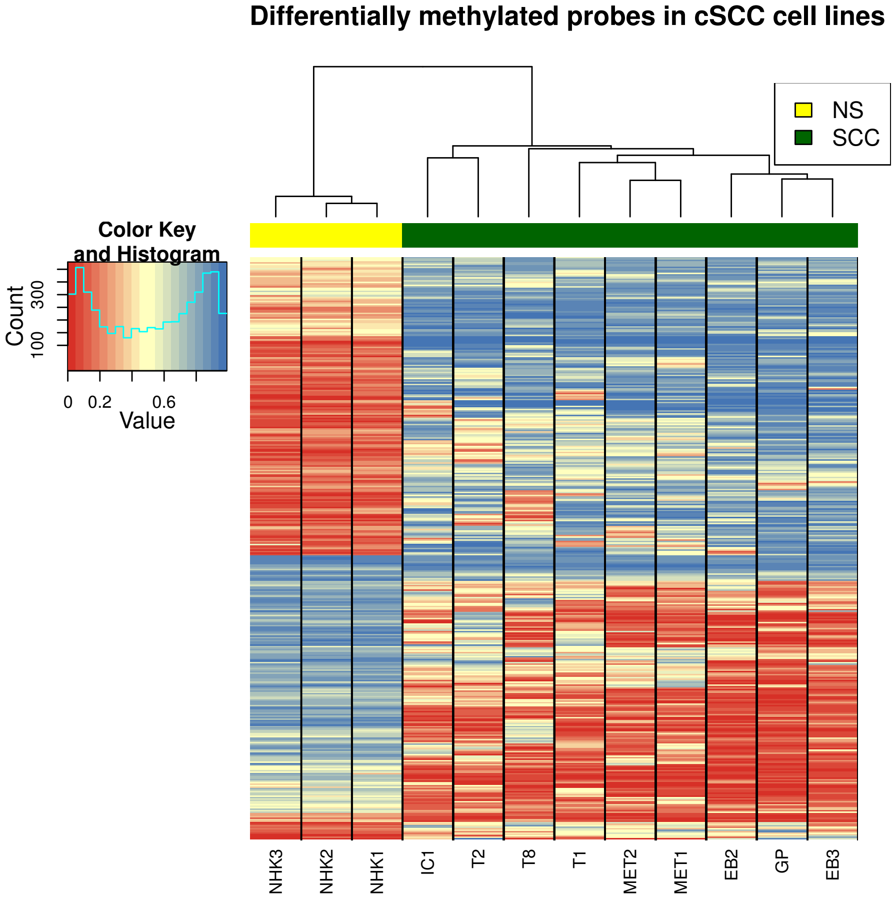


Figure 5.6 Heatmap of beta-values of probes differentially methylated in cSCC cell lines. NS=primary human keratinocytes, SCC=cSCC cell lines.

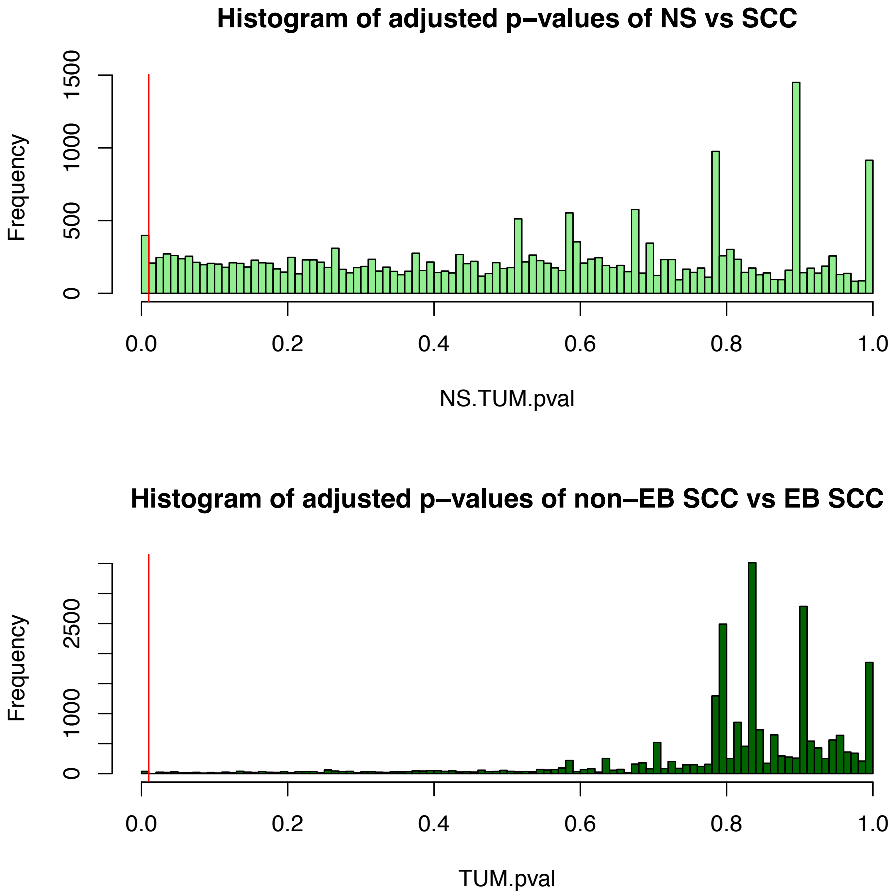


Figure 5.7 Histogram of adjusted p-values for all probes comparing PHK vs. all cSCC cell lines combined (above) and non-RDEB cSCC cell lines vs. RDEB-cSCC cell lines (below). Red line corresponds to p value of 0.01 (pre-set α level). This figure shows that a significant difference in methylation on probe-levels is far more common if all cSCC cell lines are compared with PHK than if RDEB-cSCC and non-RDEB-cSCC cell lines are compared.

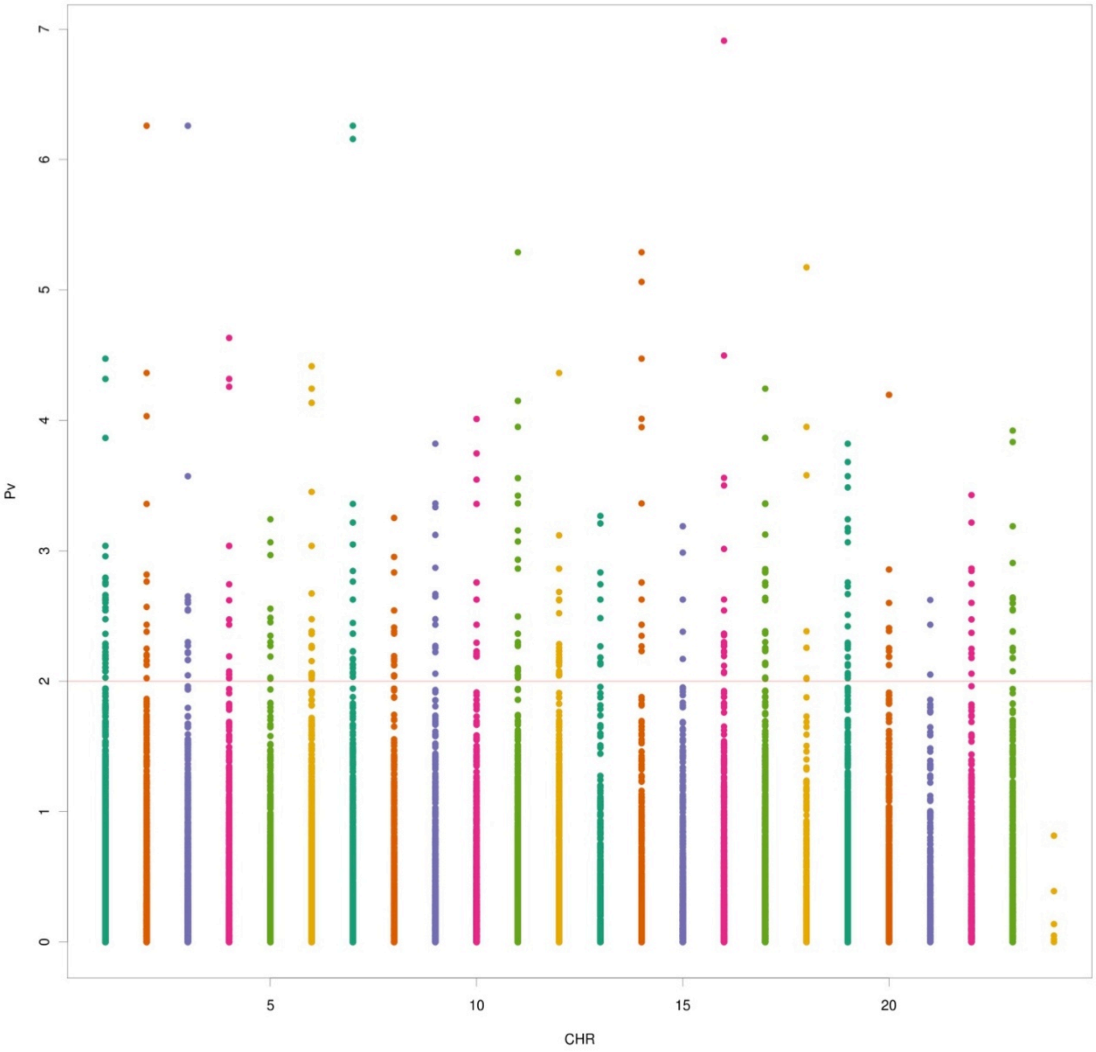


Figure 5.8 –Log10 p-values for each probe with corresponding chromosomal location. CHR=chromosome, red line corresponds to 0.01 p-value. This plot shows that differentially methylated probes are located on all chromosomes but the Y chromosome (far right, yellow). Pv=-log10 of a p-value (this converts the most significant p values to the highest integer).

This probeset represents 371 annotated genes, of which 201 (54%) were hypermethylated and 170 (46%) hypomethylated relative to normal skin. The observed skew towards hypermethylation in tumour cell lines was not statistically significant (binomial test, pvalue=0.054). Chromosomal locations of hyper- and hypomethylated genes are shown in Figure 5.9.

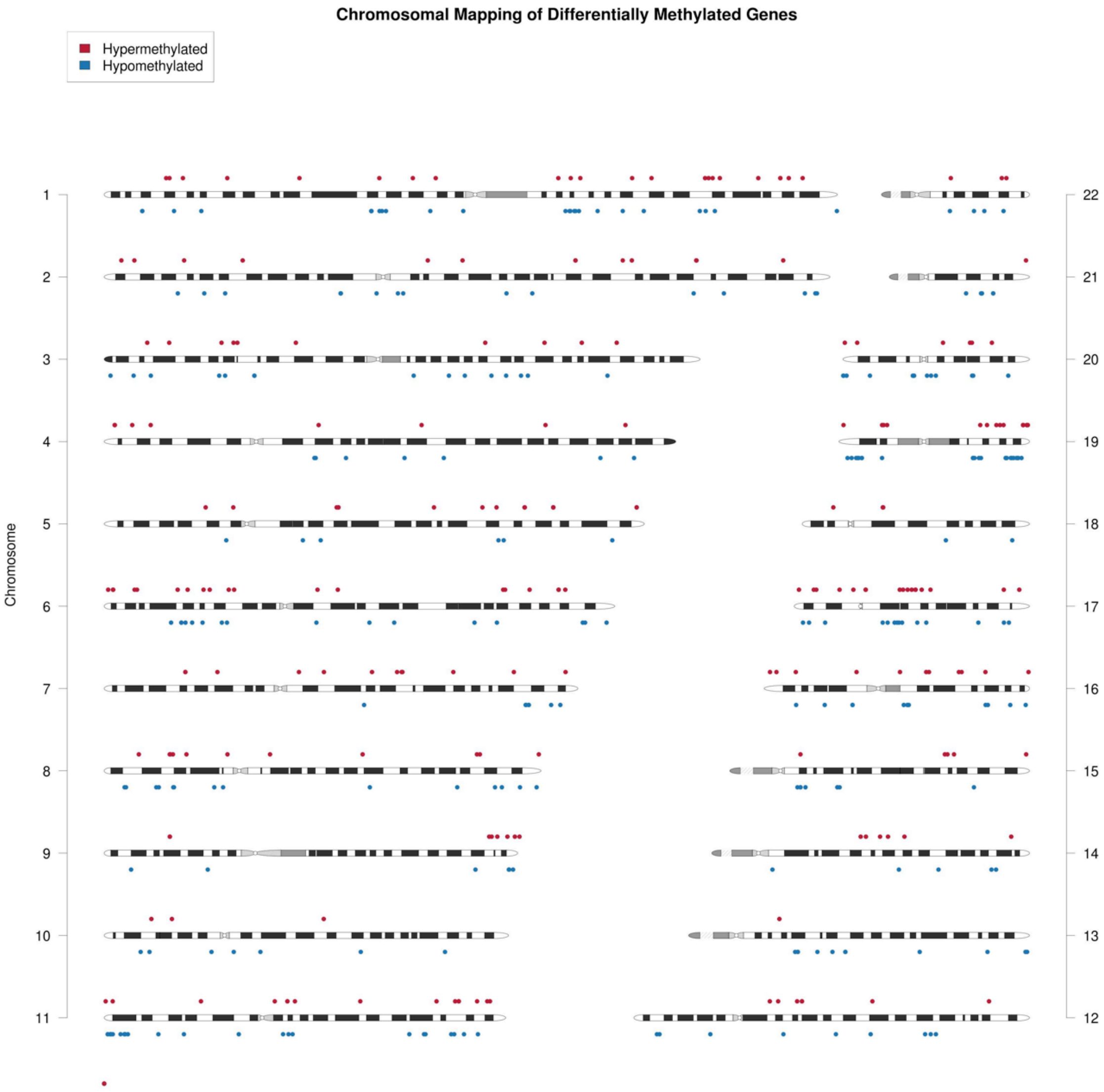


Figure 5.9 Chromosomal positions of differentially methylated genes in all cSCC cell lines.

Many of these differentially methylated genes have been previously shown to be involved in cancer, including CDKN2A, WT1, TP53I3, MRAS, GATA4, SPARCL1. Based on extensive literature search, a total of 115 (30%) genes in this set have been previously shown to be involved in cancer; either as tumour suppressor, as oncogenes or their role is ambiguous. This set of cancer-relevant genes is provided in Appendix 8.

Based on GO term annotation, many additional genes were shown to play a role in signal transduction (CHRNB4, CHRNE, ITPR2, MCHR1, F2RL3, PRLHR, GPR75, FFAR2, GABRA5, CD101, CCL8, MAPK8, IP1, NFAM1, CARD9, TAC3, DDAH2, FGD4, PLCL1, TNFRSF11B, TANK).

Four homeobox genes were identified as differentially methylated (BARHL2, PAX4, TGIF2 and POU2F2). Homeobox genes encode regulatory nuclear proteins that function as transcription factors that play a critical role during development, and dysregulation of these genes has been previously implicated in various cancers (Nunes et al., 2003).

TWIST1, a prominent inductor of EMT transition, was significantly hypomethylated in cSCC, which signifies importance of this process in this malignancy.

5.3.4. Validation of 27K array with bisulfite sequencing

DMP detected with 27K array were validated using 10 laser-capture microdissected cSCC samples and paired normal skin originating in male RTR patients. As described in Chapter 2, methylation levels of regions within 50 bp of DMP were averaged and subtracted to obtain differences in methylation.

Bisulfite sequencing data was available for 349 of 398 differentially methylated genes and the methylation levels in skin and cSCC are available in Appendix 20. The same direction of methylation difference (hyper- or hypomethylation) was confirmed for 163 genes (46.7%). 81 of the validated genes were hypo- and 82 were hypermethylated.

5.3.4.1. cSCC cell lines derived from RDEB-cSCC tumours versus cell lines derived from non-RDEB tumours

The comparison of cSCC cell lines derived from RDEB-associated tumours with non-RDEB tumours lead to the identification of 38 differentially methylated probes, which represent 37 unique genes. This geneset is listed in Table 5.3, and contains one homeobox gene (SIX6). Three genes in this set (ATP10A, GATA4, KIF5A) overlap with differentially methylated geneset derived from the comparison of normal human keratinocytes and all cSCC cell lines. 25 probes (66%) were relatively hypermethylated in RDEB-derived tumours and 13 (34%) were relatively hypomethylated. This trend towards hypermethylation in RDEB-derived tumours is statistically significant (binomial test, p-value 0.027).

Gene	CHR	Gene name	Methylation Status in RDEB-cSCC cell lines
SLC44A4	6	Solute carrier family 44, member 4	Hyper
MFRP	11	membrane frizzled-related protein	**Hypo**
HTR2C	X	5-hydroxytryptamine (serotonin) receptor 2C, G protein-coupled	Hyper
C18orf34	18	coiled-coil domain containing 178	Hyper
MGC44505	2	-	**Hypo**
SORCS3	10	sortilin-related VPS10 domain containing receptor 3	Hyper
HRC	19	histidine rich calcium binding protein	**Hypo**
SIX6	14	SIX homeobox 6	Hyper
MGC11271	19	-	Hyper
RNASE6	14	ribonuclease, RNase A family, k6	**Hypo**
PLP1	X	proteolipid protein 1	**Hypo**
GATA4	8	GATA binding protein 4	Hyper
KIF5A	12	kinesin family member 5A	**Hypo**
EYA4	6	eyes absent homolog 4	Hyper

ADRA1A	8	adrenoceptor alpha 1	Hyper
MEST	7	mesoderm specific transcript	Hyper
GALR1	18	galanin receptor 1	Hyper
VSIG4	X	V-set and immunoglobulin domain containing 4	**Hypo**
ATP10A	10	ATPase, class V, type 10A	**Hypo**
DYRK1B	19	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B	Hyper
DOK1	2	docking protein 1, 62kDa	Hyper
PRDM14	8	PR domain containing 14	Hyper
GLYAT	11	glycine-N-acyltransferase	**Hypo**
FOXG1B	14	-	Hyper
CYP2E1	10	cytochrome P450, family 2, subfamily E, polypeptide 1	Hyper
GAD2	10	glutamate decarboxylase 2	Hyper
GRIA4	11	glutamate receptor, ionotropic, AMPA 4	Hyper
GRM6	5	glutamate receptor, metabotropic 6	Hyper
FLJ42486	14	-	Hyper
C7orf9	7	-	**Hypo**
CRMP1	4	collapsin response mediator protein 1	Hyper
UNQ9433	8	-	Hyper
TCEAL7	X	transcription elongation factor A (SII)-like 7	**Hypo**
CCDC37	3	coiled-coil domain containing 37	Hyper
OR2B6	6	olfactory receptor, family 2, subfamily B, member 6	**Hypo**
DPCR1	6	diffuse panbronchiolitis critical region 1	**Hypo**
PDGFD	11	platelet derived growth factor D	Hyper

Table 5.3. DMG in RDEB-cSCC derived cell lines compared to non-RDEB cSCC derived cell lines.

Heatmap of differentially methylated probes in RDEB-cSCC cell lines is shown in Figure 5.10.

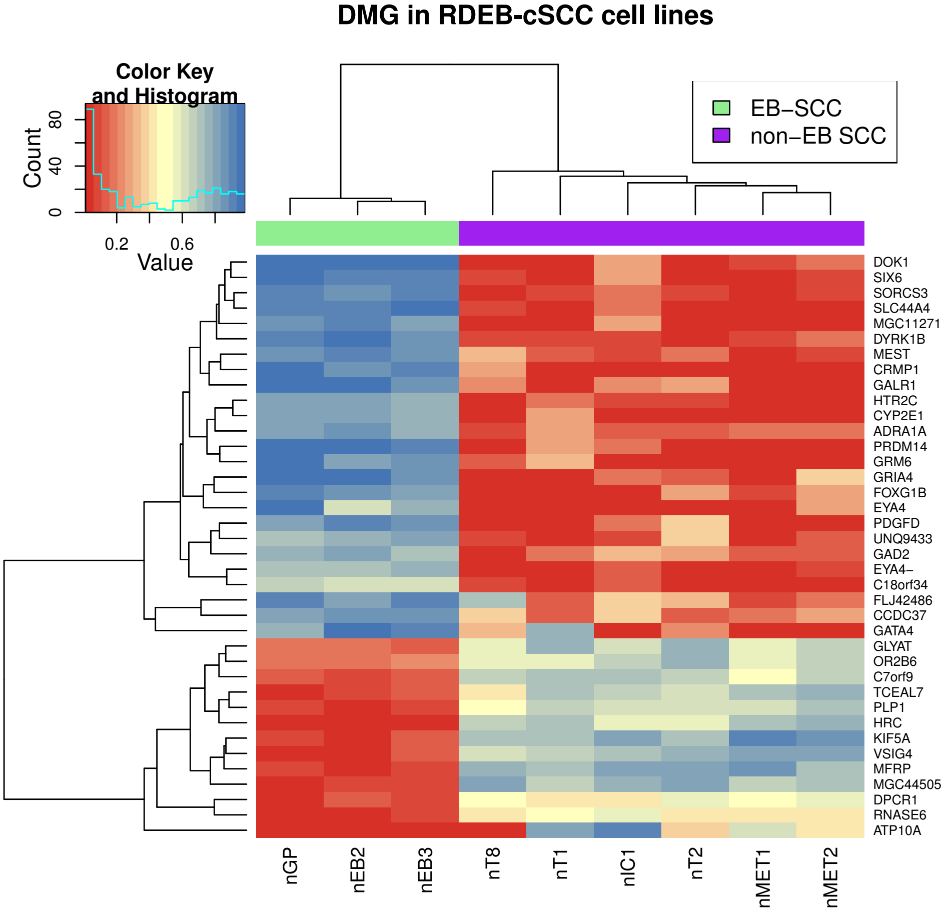


Figure 5.10 Heatmap of differentially methylated genes in RDEB-cSCC cell lines.

Based on extensive literature search, I have observed that many genes in this set were shown to be involved in mammalian development and in various disorders including psychiatric, cardiac diseases and cancer. For example, SLC44A4 has been demonstrated to be overexpressed in various epithelial cancers (Ricart, 2011), DOK1 is a putative tumour suppressor shown to be hypermethylated in a variety of cancers (Saulnier et al., 2012), and was hypermethylated in RDEB-cSCC cell lines compared to non-RDEB cSCC cell lines. The same was true for GALR1, the frequent hypermethylation of which has been detected in head and neck SCC (Misawa et al., 2008).

5.3.5. Enrichment Analysis

5.3.5.1. KEGG pathway analysis

The comparison of normal keratinocytes and cSCC cell lines methylation profiles showed that the most significant dysregulation was seen in the NOD-like receptor signalling pathway. This pathway is involved in inflammatory and apoptosis response and NF-kappaB activation, and has been implicated in gastric cancer (Hedman et al., 2002) and in NSLCC (Strazisar et al., 2009).

Dysregulated pathways directly linked to cancer included *Apoptosis* and *p53 signalling pathway*. A full list of dysregulated pathways is provided in Table 5.4.

KEGG Pathway	**Adjusted P-value**
NOD-like receptor signaling pathway	<0.0001
Apoptosis	0.0003
Maturity onset diabetes of the young	0.0012
p53 signaling pathway	0.0014
Lysine biosynthesis	0.0016
O-Glycan biosynthesis	0.0048
Sphingolipid metabolism	0.0177
Amyotrophic lateral sclerosis (ALS)	0.0265
RIG-I-like receptor signaling pathway	0.0281

Table 5.4. Significantly dysregulated pathways in differentially methylated probes in cSCC cell lines.

KEGG pathway analysis of differentially methylated probes in RDEB-cSCC cell lines compared to non-RDEB cSCC cell lines has found only 1 pathway – Neuroactive ligand-receptor activation – to be dysregulated.

5.3.6. Comparison of primary human keratinocytes and cSCC cell lines with genome-wide DNA methylation profile of human skin

To explore the similarity of PHK to normal epidermis, I used a published dataset containing 30 human skin samples hybridised to the same DNA methylation chip (Gronniger et al., 2010). As shown in Figure 5.11, PHK clustered with human skin samples obtained from healthy volunteers.

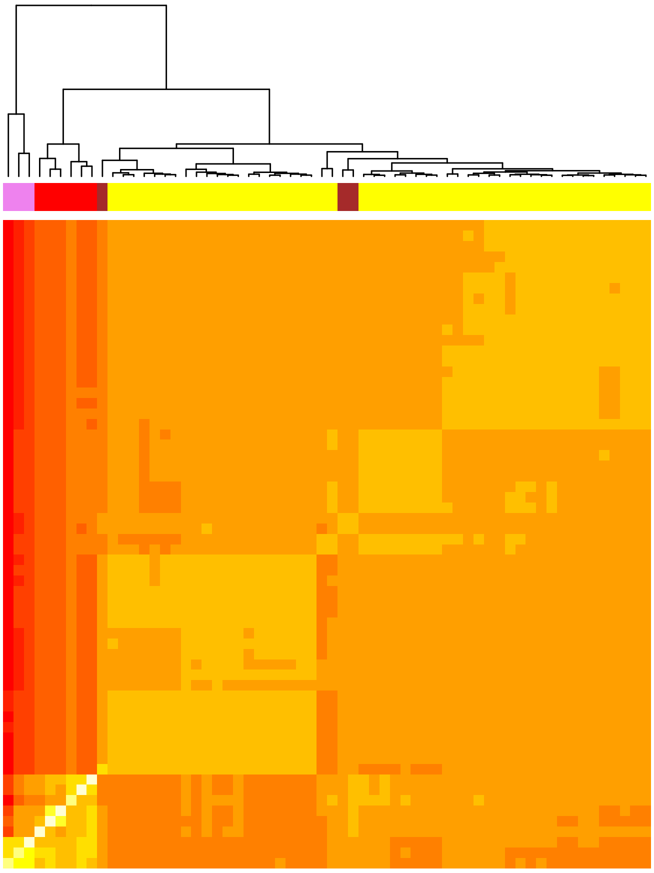


Figure 5.11 Unsupervised hierarchical clustering of genome-wide methylation profile 30 human skin samples, cultured primary keratinocytes and cSCC cell lines. Normal human skin=yellow, PHK=brown, non-RDEB cSCC cell lines=red, RDEB-cSCC cell lines=pink. This figure shows that cultured primary keratinocytes cluster with normal human skin based on Pearon's correlation of their methylomes.

To confirm the validity of 398 differentially methylated probes in cSCC cell lines, I plotted normalised methylation levels across PHK, cSCC cell lines and 30 human skin samples. As shown in Figure 5.12, methylation levels of this probeset in PHK are concordant with human skin.

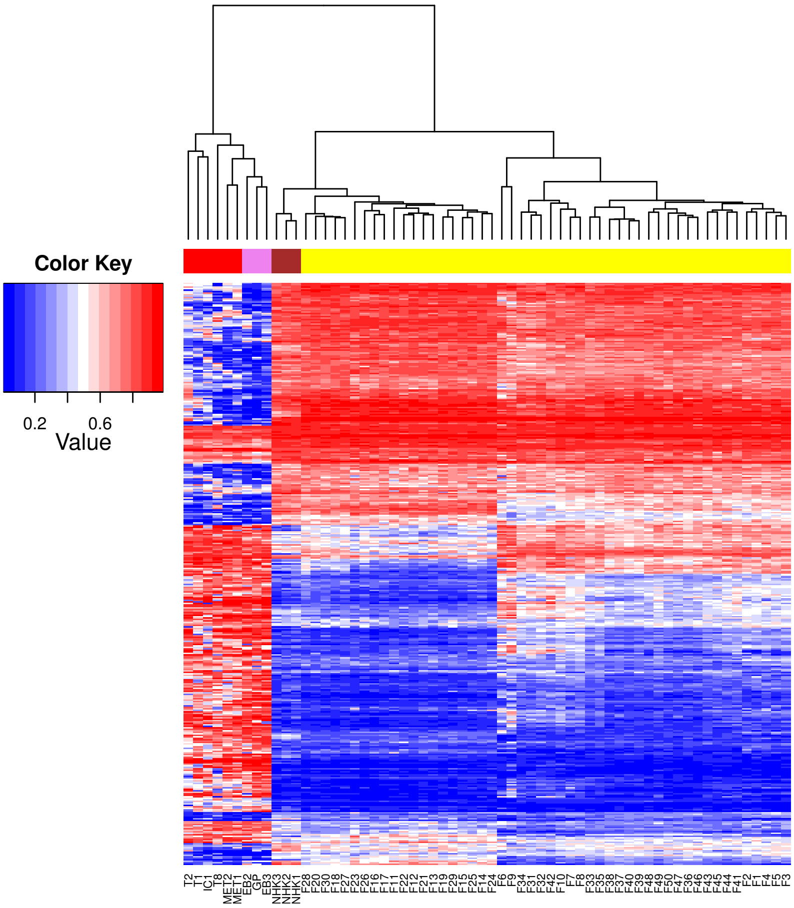


Figure 5.12 Unsupervised hierarchical clustering of 30 skin samples, PHK, non-RDEB cSCC cell lines and RDEB-cSCC cell lines based on methylation levels of 398 differentially methylated genes identified cSCC cell lines. Normal human skin=yellow, PHK=brown, non-RDEB cSCC cell lines=red, RDEB-cSCC cell lines=pink. This graph confirms consistency of PHK and normal skin methylation profile.

5.3.7. Comparison with methylation levels in additional cultured tissues

I hypothesised that due to the process of dedifferentiation in cSCC cell lines, methylation profiles of these cells would cluster with less differentiated cell types. To test this hypothesis, I used methylation profiles derived from a panel of tissues hybridised to the same kind of microarray chip, and explored the clustering patters of genome-wide methylation profiles of this dataset with cSCC cell lines following standard data manipulation (normalisation, data filtering).

As shown in Figure 5.13, methylation profile of cSCC cell lines and hESC and iPSC is quite distinct from the remaining tissues. Additionally, PHK clustered with MSC, fetal lung fibroblasts and cord-blood endothelial precursors, rather than forming a distinct cluster. This suggests that dedifferentiation in cancer cells occurs due to the gain of unique molecular properties rather than gaining global, genome-wide methylation profile similar to other less differentiated or completely undifferentiated tissues.

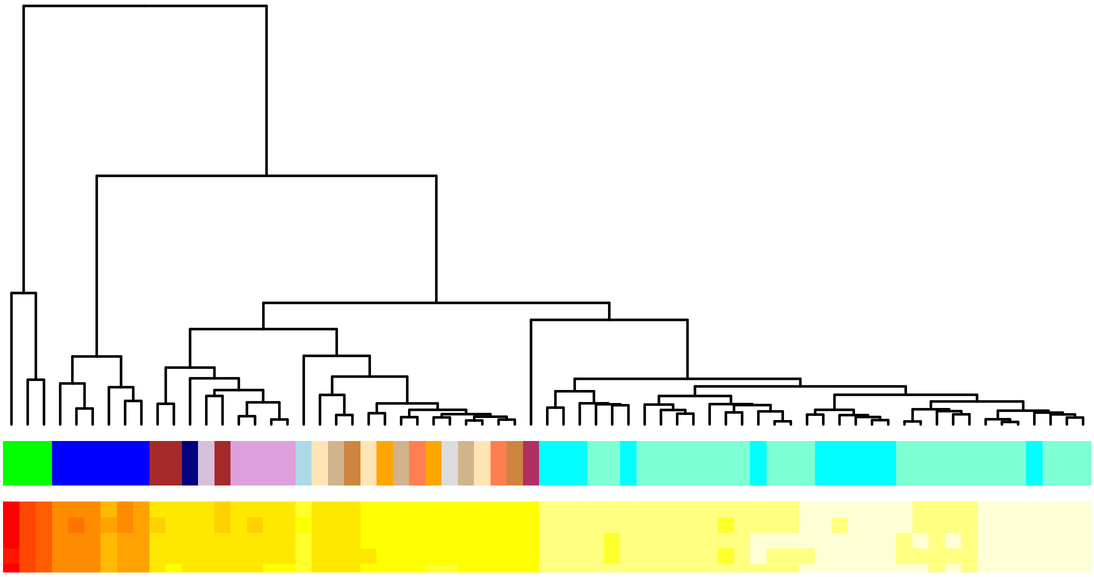


Figure 5.13 Clustering of genome-wide methylation profiles of various tissues. cSCC cell lines (green=RDEB-cSCC, blue=non-RDEB cSCC) cluster independently of this dataset, while PHK (brown) form a cluster with cord-blood endothelial progenitor cells (light pink) and fetal lung fibroblasts (navy blue), and a greater cluster with MSC (pink). iPSC and hESC form a distinct cluster on the right (light blue and aquamarine), and remaining samples (cord blood CD34+ cells, peripheral blood CD34+ cells, bone marrow CD34+ cells, peripheral blood mononuclear cells, cord blood mononuclear cells, human umbilical vein endothelial cells, cord blood CD34- cells) form a cluster in the middle.

I then looked at average differences in methylation between selected individual tissue types by averaging beta-values for each probe in tissue type, and then calculating average of those differences: Primary human keratinocytes, mesenchymal stem cells, fetal lung fibroblasts, cord blood endothelial progenitor cells, cord blood CD34+ cells, peripheral blood CD34+ cells, bone marrow CD34+ cells, hESC, iPSC, and human umbilical vein endothelial cells. As shown in Figure 5.14, cSCC cell lines showed the highest differences in comparison with any additional tissue type, suggesting that the methylation profile of cancer cells is entirely unique.

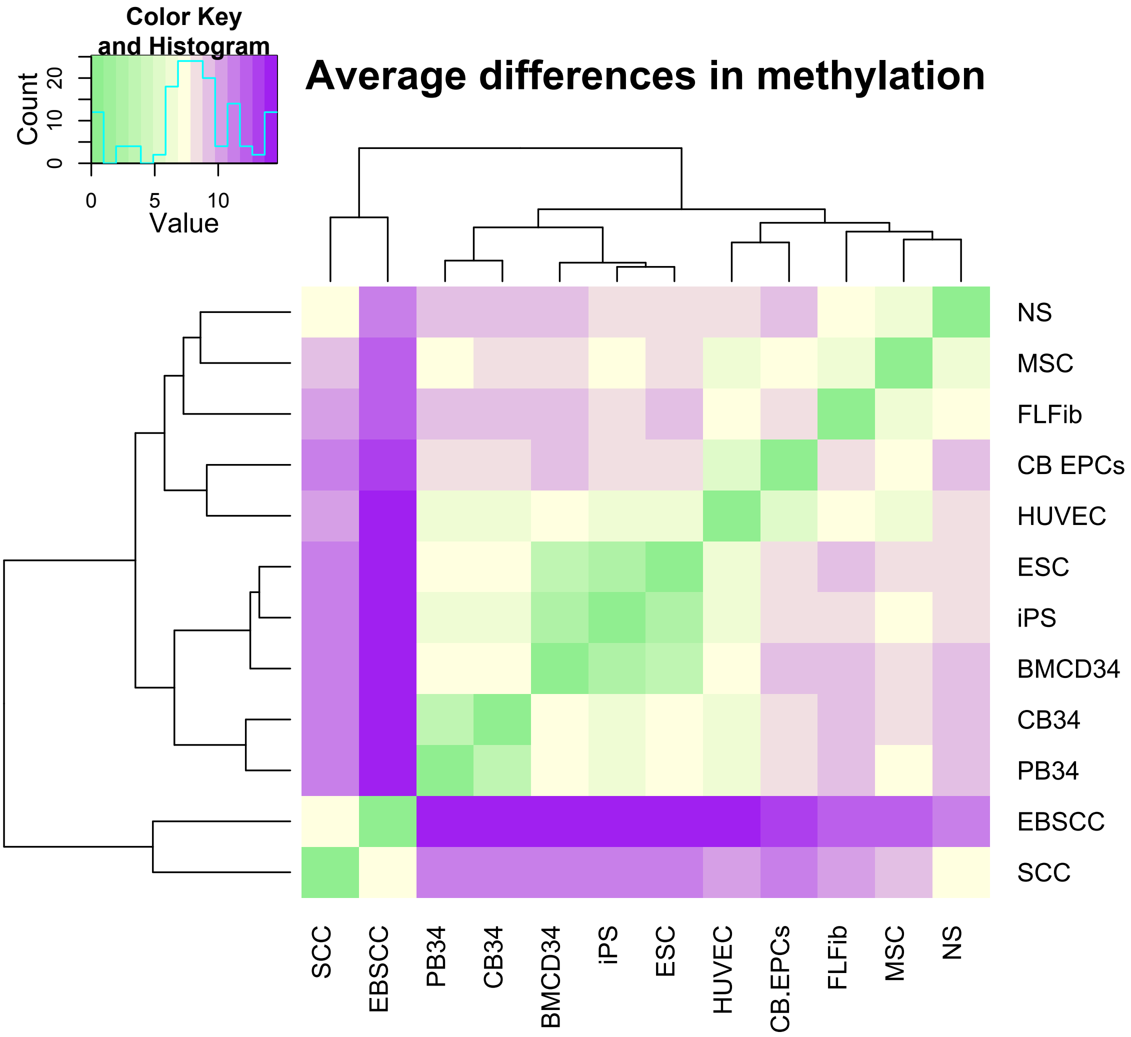


Figure 5.14 Average differences in methylation between all tissue types in percentage points. Each colour in the colour key corresponds to a difference in beta-values in percentage points (0=0%, 5=5%, 10=10%). The highest differences are between cancer cell lines (SCC=cSCC, EBSCC=RDEB cSCC) and the remaining tissues. Conversely, hESC, iPS and bone marrow-derived CD34+ progenitor cells (BMCD34) show on average 5% differences. Cultured normal keratinocytes (NS) show the lowest average differences in comparison with mesenchymal stem cells (MSC) and also fetal lung fibroblasts (FLFib). PB34=peripheral blood-derived CD34 positive cells, CB34=cord blood CD34+ cells, HUVEC=human umbilical vein endothelial cells, CB EPCs=cord blood endothelial progenitor cells.

5.3.7.1. Human embryonic stem cells signature in cSCC cell lines

I then plotted the 398 differentially methylated probes in PHK, cSCC cell lines and hESC and iPSC cells. Figure 5.15 shows that while the majority of differentially methylated probes in cSCC showed consistent levels of methylation in hESC and iPSC, a small subset of probes (highlighted in black rectangles) showed methylation levels that were distinct in PHK yet showed practically equal methylation levels across cSCC and iPSC/hESC cells.

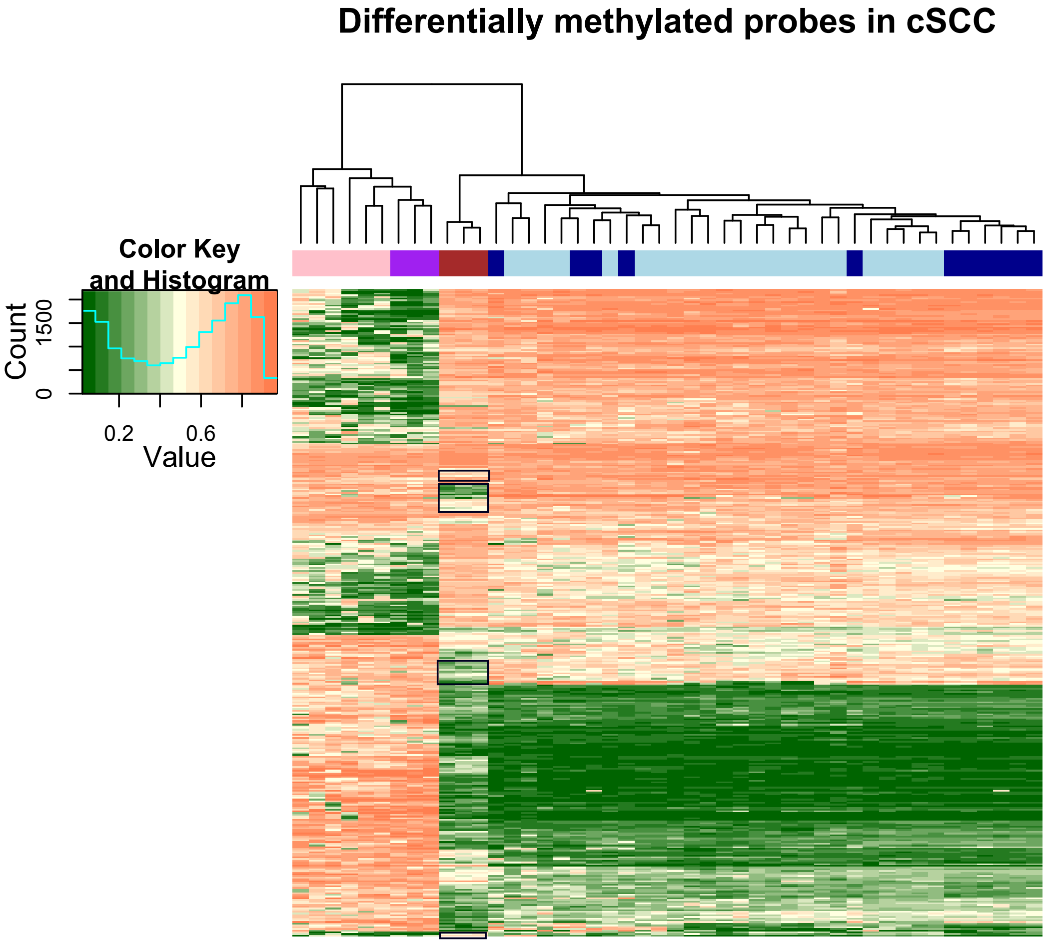


Figure 5.15 398 differentially methylated probes in PHK (dark red), cSCC cell lines (pink), RDEB-cSCC cell lines (purple), hESC (dark blue) and iPSC (light blue). Black rectangles indicate probes that show similar methylation levels in cSCC cell lines and hESC/iPSC, yet are differentially methylated in PHK.

I then explored the methylation levels of those probes closer by plotting only cSCC cell line and iPSC/hESC probe data. In order to visually explore the differences at more subtle levels, I used additional colours for distinct methylation values. As shown in Figure 5.16, a small subset of probes seemed to show consistent levels of methylation across cSCC and iPSC/hESC.

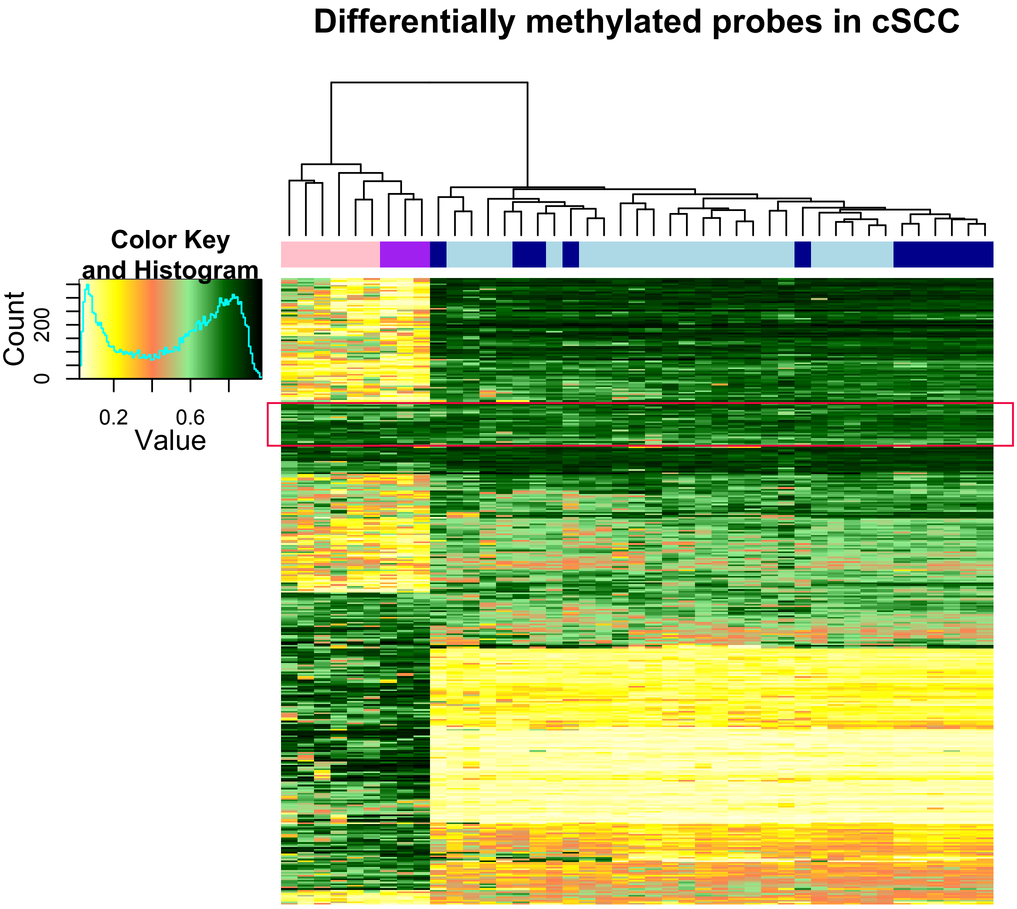


Figure 5.16 Differentially methylated probes in cSCC with corresponding levels in iPSC (light blue) and hESC (dark blue). Red rectangle highlights probes that seem to have equivalent methylation levels in those samples.

I then identified probes that show less than 5% difference in methylation levels in hESC compared to cSCC cell lines, and more than 20% in hESC compared to PHK (details are described in Chapter 5.2.4.1.). This approach resulted in 21 probles (individual methylation levels are shown in Figure 5.17) that translated in 20 genes listed in Table 5.5 to be identified as "hESC signature" in cSCC cell lines.

Gene **and** Gene name	Comment	Reference	Status in cSCC/hESC
ACTA2 **smooth muscle alpha-actin**	Marker of muscle differentiation, p53 target	(Guan et al., 2006, Botcheva et al., 2011)	Hypermethylated
AQP9 **aquaporin 9**	Upregulated in mouse ESC by MAPK as a stress response	(Bell et al., 2009)	Hypermethylated
BCDIN3 **methylphosphate capping enzyme**	Represses translation; vertebrate development; methylates the 5' end of 7SK RNA	(Singh et al., 2011)	Hypermethylated

CDKN2A **[P16] cyclin-dependent kinase inhibitor 2A**	Tumour suppressor, unmethylated in hESC	(Ohm et al., 2007)	**Hypomethylated**
DSG1 **desmoglein1**	Supports epidermal differentiation	(Harmon et al., 2013)	Hypermethylated
EDNRB **endotelin B receptor**	Hypermethylated in various cancers	(Chen et al., 2013, Watanabe et al., 2014)	Hypermethylated
F2RL3 **Protease-activated receptor 4 (PAR4), previously known as Coagulation factor II (thrombin) receptor-like 3**	Decreased expression in human gastric cancer through hypermehtylation	(Zhang et al., 2011)	Hypermethylated
FAM20B	Expressed in haematopoietic cells	(Nalbant et al., 2005)	**Hypomethylated**
FAM50B	Imprinted gene	(Nakabayashi et al., 2011)	Hypermethylated
GGT1 **Human gamma-glutamyltransferase**	Expression is characteristic of hepatoblasts	(Cipak et al., 2010)	Hypermethylated
KIAA0753	Underexpressed in non-adherent cancer cells	(Marquez et al., 2013)	**Hypomethylated**
KRTAP1-1 **keratin associated protein 1-1**	Expressed in hair follicle cortex	(Shimomura et al., 2002)	Hypermethylated
PCDHGA12 **protocadherin gamma subfamily A, 12**	Methylated in leukaemia	(Taylor et al., 2007)	Hypermethylated
PPAP2B **phosphatidic acid phosphatase type 2B**	Involved in tumour regulation through β-catenin and CYCLIN-D1 signaling.	(Chatterjee et al., 2011)	Hypermethylated
SAA2 **serum amyloid A2**			Hypermethylated
SERPINA3 **serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3**	Hypermethylated in hESC	(Kim et al., 2011)	Hypermethylated
SLC22A18 **solute carrier family 22, member 18**	Paternally imprinted, putative tumour suppressor	(Jung et al., 2011, Chu et al., 2011)	Hypermethylated
SLC43A3 **solute carrier family 43, member 3**	Low expression in hESC	(Zhang et al., 2007, Ulloa-Montoya et al., 2007)	Hypermethylated
TSPAN16	Decreased expression in a	(Juric et al., 2007)	Hypermethylated

tetraspanin 16	subtype of acute lymphoblastic leukaemia	
ZNF662	NA	Hypermethylated

Table 5.5 Genes that represent hESC signature in cSCC cell lines based on methylation levels.

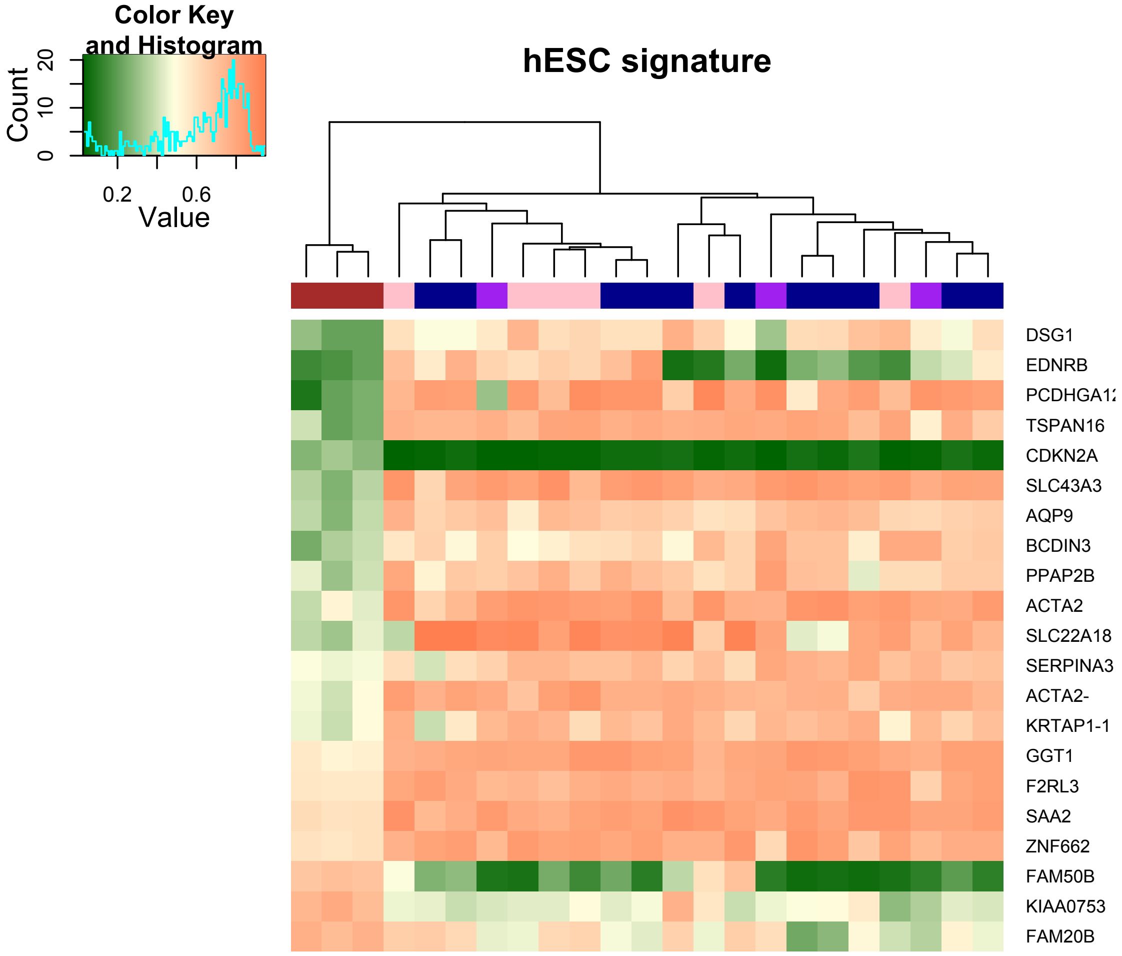


Figure 5.17 Probes that represent hESC signature in cSCC cell lines. Red samples on the left=PHK, pink samples=non-RDEB cSCC cell lines, purple samples=RDEB cSCC cell lines, dark blue samples=hESC.

5.3.8. Transcriptional profiling of cSCC cell lines

I further explored the impact of methylation changes on transcriptional regulation in six primary keratinocytes (PHK), three hESC cell lines and previously described cSCC cell lines. One PHK sample has not passed the quality control and was excluded from further study.

As shown in Figure 5.18, hESC possess a very distinct transcriptional profile. This figure also shows that PHK are transcriptionally distinct from cSCC cell lines, and warranted comparative detection of differentially transcribed genes. If PHK clustered within cSCC cell lines, this would indicate that their transcriptional profile is not likely to be significantly different from cSCC cell lines. Unlike in clustering analysis of methylation profiles, transcriptional profiling of cSCC cell lines did not segregate based on original cSCC histology (data not shown).

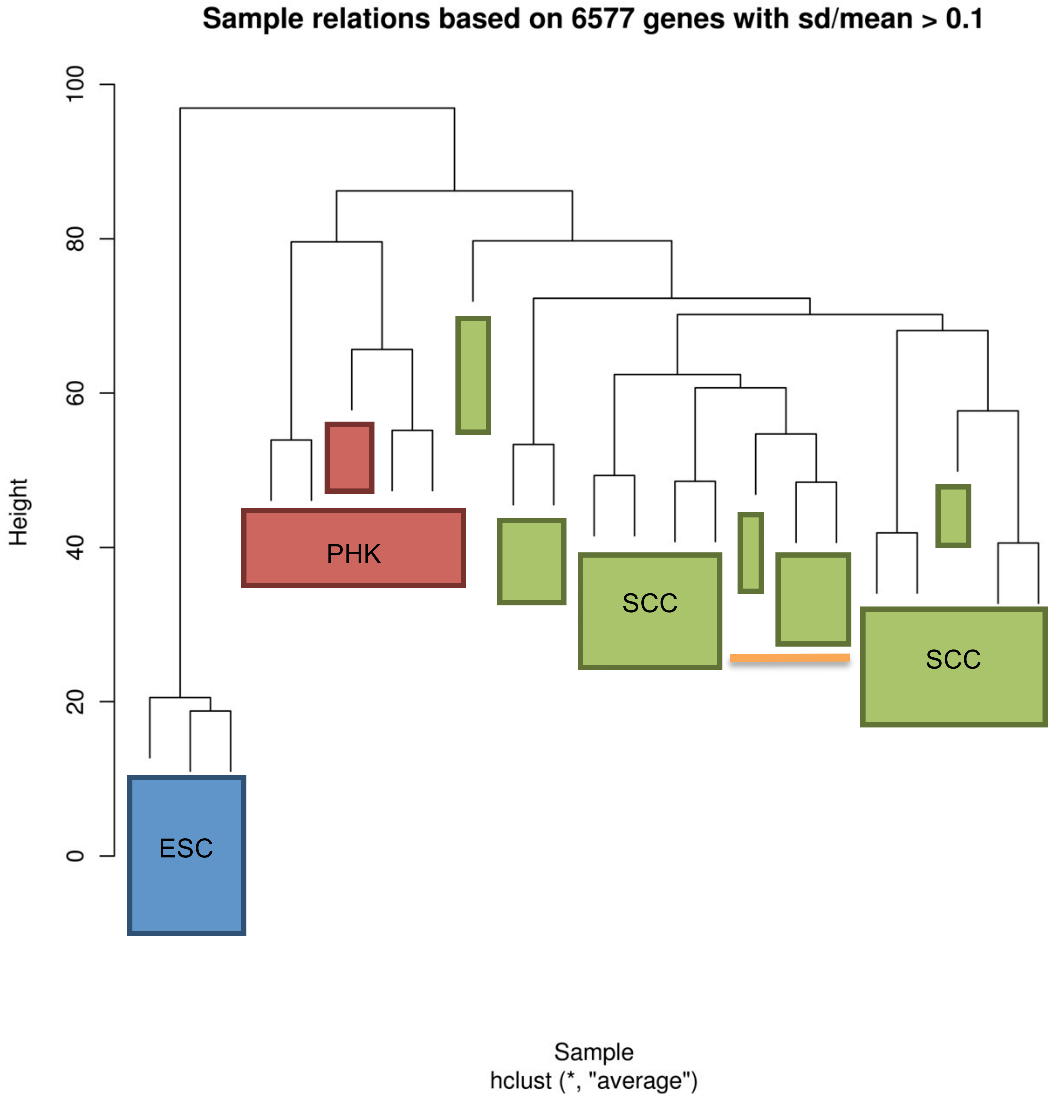


Figure 5.18 Clustering of transcriptional profiles confirms correlation between samples of the same tissue origin. Blue=hESC, green=cSCC cell lines, red=PHK. Orange line indicates a cluster of three RDEB-cSCC cell lines.

5.3.8.1. Differentially transcribed genes in cSCC cell lines

I applied Bayesian statistics to determine differentially expressed probes comparing cSCC cell lines and primary human keratinocytes. This probabilistic approach uses various parameters, to compensate for the lack of replication in microarray data, and has been demonstrated as superior to simple fold-change or non-parametric analysis of microarray expression data (Tang et al., 2013). Table 5.6 summarizes the results in terms of number of differentially expressed probes using various thresholds; Table 5.7 lists the top 31 differentially expressed probes.

Adjusted pvalue threshold	Number of Differentially Expressed probes
<0.0001	31
<0.001	199
<0.01	991
<0.05	2510

Table 5.6 Differentially expressed probes comparing cSCC cell lines and primary human keratinocytes using various cut-off thresholds.

geneSymbol	geneName	logFC	AveExpr	adj.P.Val
ITM2A	integral membrane protein 2A	-3.598319171	7.199052039	<0.0001
INPP5D	inositol polyphosphate-5-phosphatase, 145kDa	-2.086989877	6.421191649	<0.0001
COL22A1	collagen, type XXII, alpha 1	-1.697063889	6.358206802	<0.0001
SPATA18	spermatogenesis associated 18 homolog (rat)	-1.365464501	6.290337707	<0.0001
BNIPL	BCL2/adenovirus E1B 19kD interacting protein like	-3.190986731	7.239378147	<0.0001
VWA5A	von Willebrand factor A domain containing 5A	-1.99319546	6.953819069	<0.0001
C9orf150	chromosome 9 open reading frame 150	-1.730644629	6.706085831	<0.0001
TWIST1	twist homolog 1 (Drosophila)	1.858724861	7.461149001	<0.0001
WFDC5	WAP four-disulfide core domain 5	-2.120210188	6.70350485	<0.0001
SLC20A2	solute carrier family 20 (phosphate transporter), member 2	-1.398189952	8.628350267	<0.0001
C1orf59	chromosome 1 open reading frame 59	2.562296627	8.709763244	<0.0001
ATP12A	ATPase, H+/K+ transporting, nongastric, alpha polypeptide	-2.68372915	6.702343095	<0.0001
NISCH	nischarin	-1.271723362	7.725084387	<0.0001
TRAK1	trafficking protein, kinesin binding 1	-1.835421466	8.612845799	<0.0001
COL18A1	collagen, type XVIII, alpha 1	-1.776459729	6.432226465	<0.0001
FYCO1	FYVE and coiled-coil domain containing 1	-1.356963286	7.534225602	<0.0001
BTG2	BTG family, member 2	-2.09909291	7.659728098	<0.0001
HSPC159	galectin-related protein	-2.331471633	8.446473019	<0.0001
LRRFIP2	leucine rich repeat (in FLII) interacting protein 2	-1.683209918	9.195168392	<0.0001
C6orf150	chromosome 6 open reading frame 150	2.169262015	7.314817213	<0.0001
ANKRD35	ankyrin repeat domain 35	-1.780236775	6.427049067	<0.0001
HOXB7	homeobox B7	2.811943742	8.033126094	<0.0001
CDSN	corneodesmosin	-0.919662161	6.107765422	<0.0001
UPK1B	uroplakin 1B	-1.059026093	6.223324642	<0.0001
PPFIBP2	PTPRF interacting protein, binding protein 2	-2.217955791	6.778286212	<0.0001

	(liprin beta 2)			
SESN1	sestrin 1	-1.050584103	7.058372163	<0.0001
LY6D	lymphocyte antigen 6 complex, locus D	-2.083147479	6.223821922	<0.0001
C7orf10	chromosome 7 open reading frame 10	-2.156236189	7.871272618	<0.0001
NA	NA	-0.830028562	6.53044291	<0.0001
NA	NA	-0.690358656	6.518646325	<0.0001
GGT6	gamma-glutamyltransferase 6	-1.475120675	6.239338224	<0.0001

Table 5.7 Most differentially expressed probes in cSCC cell lines compared to primary human keratinocytes. LogFC indicates down- or up-regulation in cSCC cell lines; negative sign means down-regulation. Most genes on the list are in fact under-expressed in cSCC cell lines.

Heatmap of this probeset is shown in Figure 5.19.

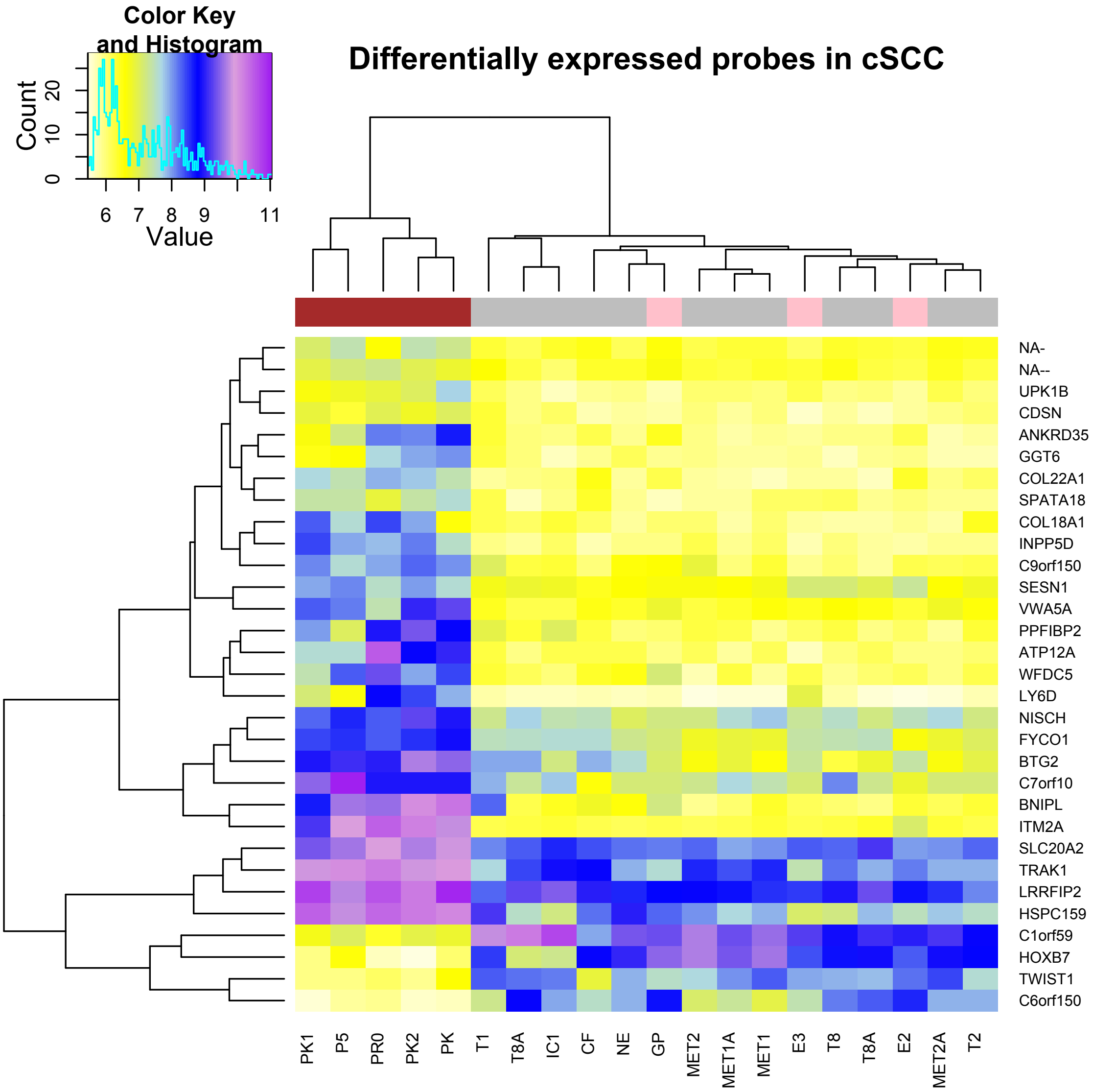


Figure 5.19 Heatmap of top 31 differentially expressed probes in cSCC cell lines. PHK=brown, RDEB cSCC lines = pink, non-RDEB cSCC lines=grey.

5.3.8.2. Comparison of expression profile with methylation data

To evaluate correlation between methylation and transcription changes, we compared the list of differentially methylated genes with differentially expressed genes to assess potential overlap of those genelists.

The comparison of probes with pvalue <0.01 in both lists has discovered only 16 overlapping genes (AQP9, CCT6A, DSG1, GABRP, GSN, IVL, KYNU, MACF1, NA, NDRG2, PRKAR1B, SERPINA3, SLC43A3, TRIM2, TWIST1, ZNF667), using a less stringent pvalue of <0.05, 95 genes were found to be present in both lists (95/1384, 6.8% of differentially methylated genes, 95/2474, 3.8% of differentially expressed genes). This finding confirms previously shown data suggesting that the highest differences in methylation do not necessarily translate linearly into corresponding transcription changes.

To further explore the relationship between methylation and expression changes, methylation and transcription levels in PHK were paired based on gene ID (entrez ID), and plotted against one another. Given that methylation levels gain values from 0 to 1 and transcription levels range from 6 to 15 after log2-transformation of raw values detected on the array, in order to make the two dimensions comparable, transcription values were divided by the highest value, thus the new maximum is 1. This operation made the two dimensions easier to compare in a single plot.

As shown in Figure 5.20, after segregating the methylation and corresponding expression values in bins of <1% methylation, 1-5%, 5-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, 90-95%, 95-100% methylation, expression levels generally tend to decrease as methylation increases, but increase again in bins with the highest methylation levels. A greater detail of this trend is depicted in Figure 5.21.

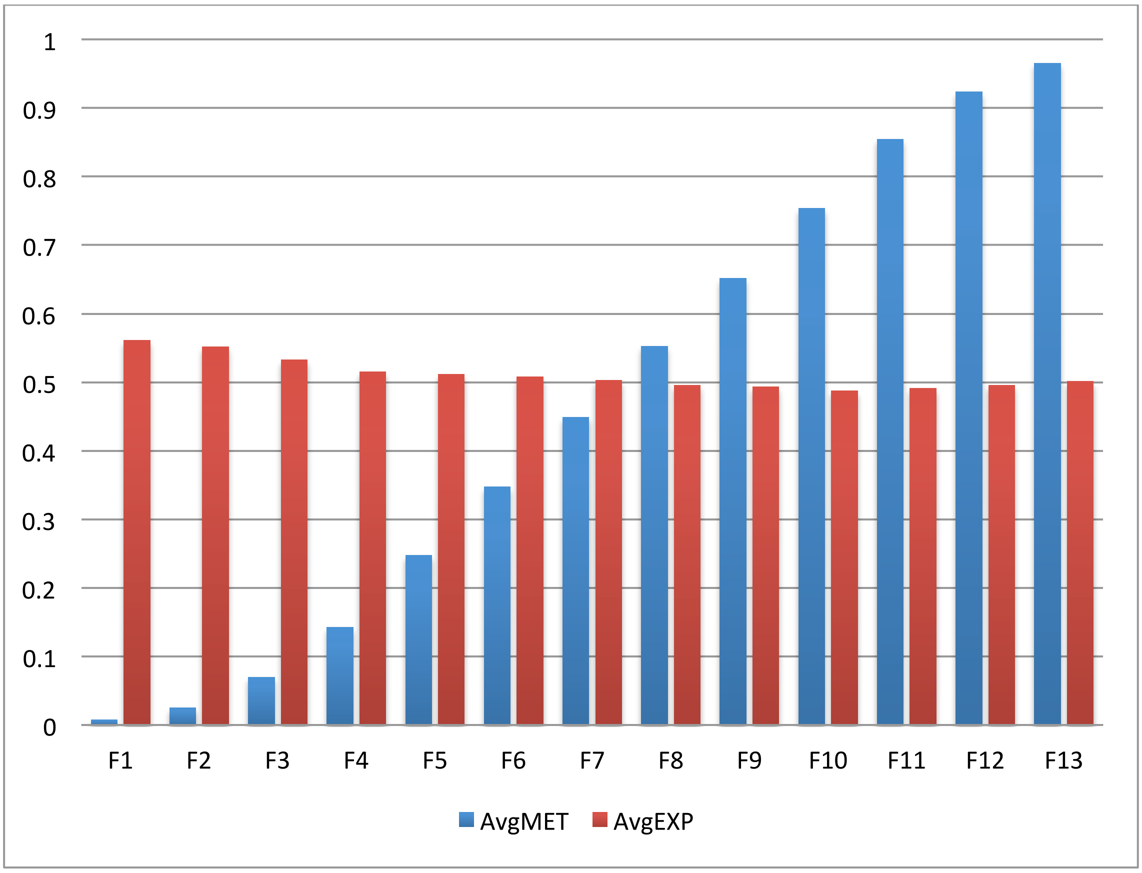


Figure 5.20 Methylation and expression levels in PHK. As methylation increases, expression generally decreases, but with the highest methylation levels, expression levels increase. AvgMET=average methylation, AvgEXP=average expression levels. F1=<1%, F2=1-5%, F3=5-10%, F4=10-20%, F5=20-30%, F6=30-40%, F7=40-50%, F8=50-60%, F9=60-70%, F10=70-80%, F11=80-90%, F12=90-95%, F13=95-100%.

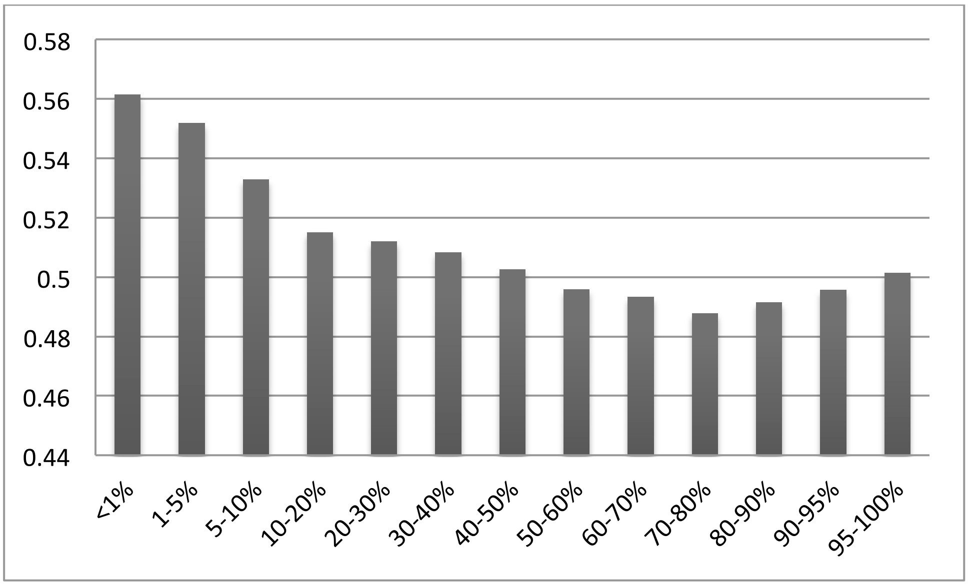


Figure 5.21 Average expression levels of methylation bins. Methylation levels indicated along the x-axis. Bimodal pattern of expression levels apparent in this graph.

Moreover, this trend does not reflect general distribution of expression levels in the samples shown as histogram of log2 transformed values in Figure 5.22 and as a cumulative density plot in Figure 5.23.

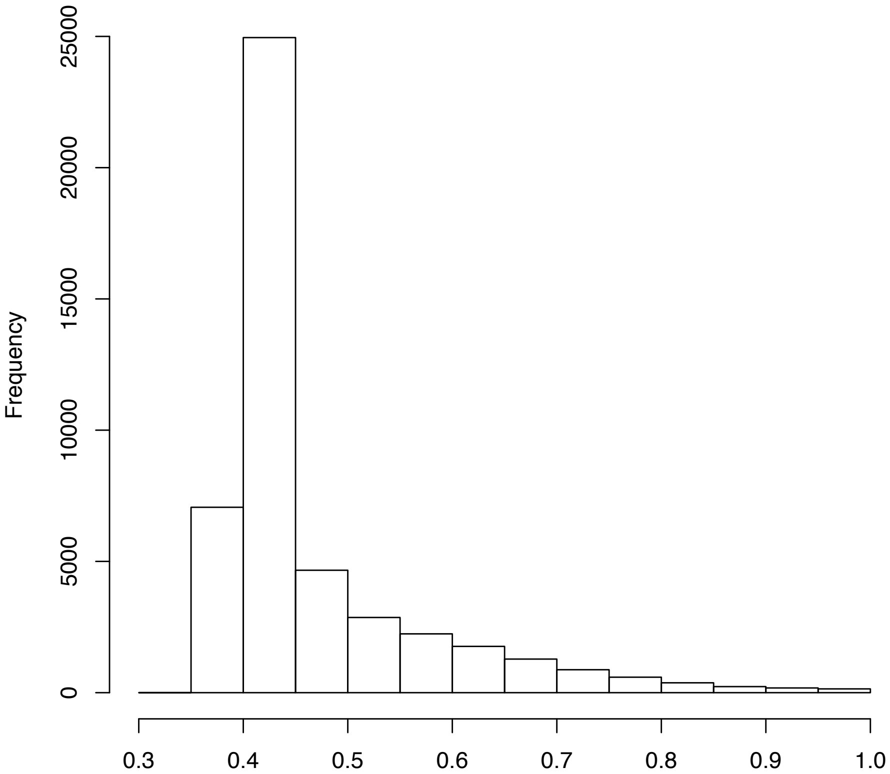


Figure 5.22 Histogram of expression values in a representative sample. X axis = expression values. This figure shows that the majority of expression values was between 0.4-0.45, thereby

the bimodal pattern observed in Figure 5.21 is not a simple repetition of general distribution of expression levels in the sample.

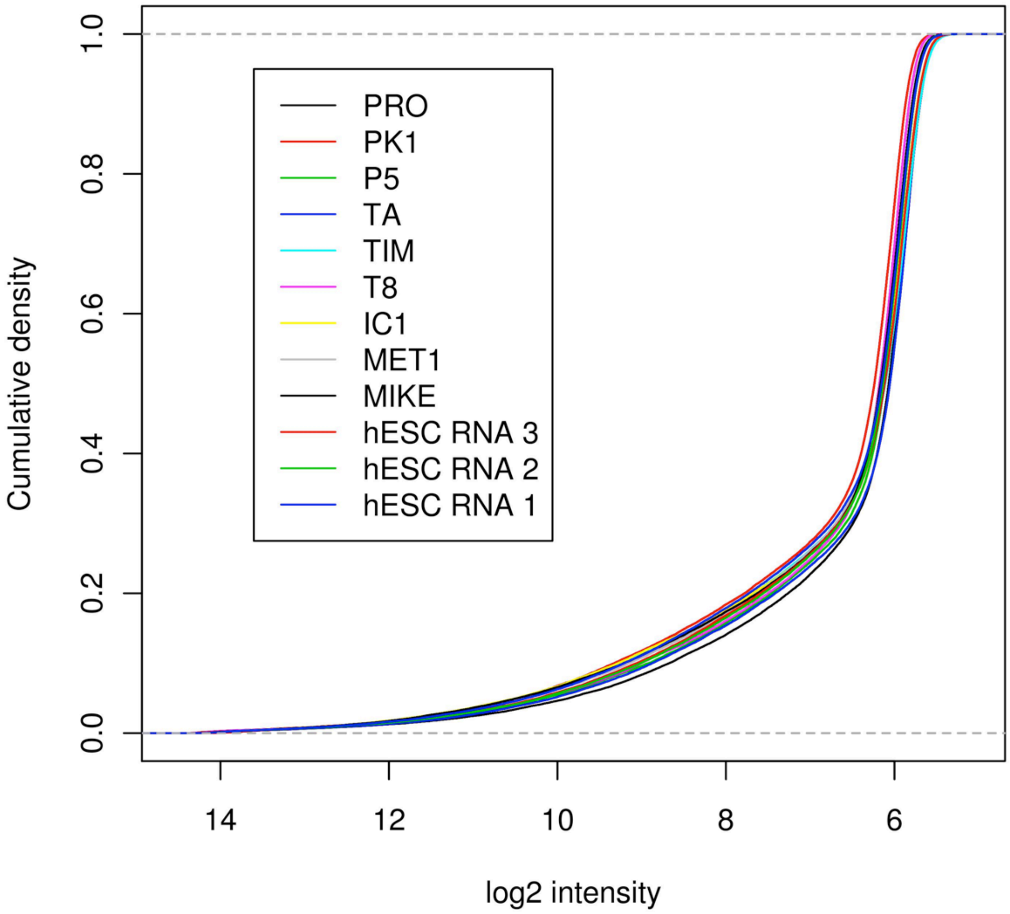


Figure 5.23 Cumulative density of log2 intensities of 12 samples on an Illumina HT12 v4 chip. Similar to the previous figure, this graph shows that the majority of genes are expressed with log2 transformed intensity of 6 or less (low level of expression), and that there is no sudden peak of high expression by a skew due to the array design as in the methylation array.

However, this trend requires further exploration as this „bimodal peak“ may be due to several factors such as distance from transcription site, the absolute number of genes which cluster into given methylation status and other factors.

5.3.8.3. KEGG pathway and GO term analysis of differentially expressed and differentially methylated genes

Given the expected absence of correlation in methylation and expression levels on individual probe level, I then hypothesised that certain critical oncogenic mechanisms may be perturbed at the DNA and mRNA level in independent genes that are involved in shared molecular processess.

To test this hypothesis, I annotated differentially methylated and differentially expressed probes (p value <0.01) with their KEGG pathway and GO terms, and identified overlapping pathways and GO terms among those detected as enriched in those datasets.

KEGG pathway analysis in differentially methylated genes is described in 5.3.5.1. In differentially expressed genes, this analysis has revealed a total of 28 pathways listed in Table 5.8.

Description	Adjusted p-value
Cell cycle	6.35E-05
DNA replication	1.04E-03
Bladder cancer	1.74E-03
Ribosome	1.81E-03
Purine metabolism	3.88E-03
Prion diseases	4.12E-03
Progesterone-mediated oocyte maturation	6.30E-03
Pyrimidine metabolism	1.23E-02
Peroxisome	1.83E-02
Pathways in cancer	1.83E-02
Oocyte meiosis	2.18E-02
RNA degradation	2.40E-02
Prostate cancer	2.42E-02
Lysosome	2.42E-02
Retinol metabolism	2.76E-02
Spliceosome	2.91E-02
p53 signaling pathway	3.34E-02
Tight junction	3.34E-02
Melanoma	3.43E-02
Ubiquitin mediated proteolysis	3.53E-02
Adherens junction	3.77E-02
Notch signaling pathway	3.77E-02
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	3.77E-02
Proteasome	3.77E-02
MAPK signaling pathway	4.01E-02
Drug metabolism - other enzymes	4.32E-02
Vascular smooth muscle contraction	4.32E-02
Non-small cell lung cancer	4.92E-02

Table 5.8 KEGG pathways enriched in differentially expressed genes (<0.01 p value).

A total of 4 pathways overlapped between the two sets: Cell cycle, Oocyte meiosis, p53 signalling pathway and Pathways in cancer.

Genes listed in Table 5.9 were involved in each pathway in the datasets. No genes disrupted at methylation and transcriptional level in each pathway overlap.

Pathway	DifEXP genes	Gene name	DifMET genes	Gene name
Cell cycle	MCM2	MCM2 minichromosome maintenance deficient 2, mitotin (S. cerevisiae)	ATM	ataxia telangiectasia mutated (includes complementation groups A, C and D)
	MCM4	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)	CDK2	cyclin-dependent kinase 2
	MCM5	MCM5 minichromosome maintenance deficient 5, cell division cycle 46 (S. cerevisiae)	CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	HDAC1	histone deacetylase 1
	CDK4	cyclin-dependent kinase 4	PRKDC	protein kinase, DNA-activated, catalytic polypeptide
	CDC26	cell division cycle 26	/	/
	PLK1	polo-like kinase 1 (Drosophila)	/	/
	CDC25B	cell division cycle 25B	/	/
	PTTG1	pituitary tumor-transforming 1	/	/
	GADD45A	growth arrest and DNA-damage-inducible, alpha	/	/
	BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	/	/
	RAD21	RAD21 homolog (S. pombe)	/	/
	POLE4	polymerase (DNA-directed), epsilon 4 (p12 subunit)	/	/
Oocyte meiosis	CDC26	cell division cycle 26	PPP3R2	protein phosphatase 3 (formerly 2B), regulatory subunit B, 19kDa, beta isoform (calcineurin B, type II)
	PLK1	polo-like kinase 1 (Drosophila)	CDK2	cyclin-dependent kinase 2
	PTTG1	pituitary tumor-transforming 1	INS	insulin
	PRKX	protein kinase, X-linked	PPP1CC	protein phosphatase 1, catalytic subunit, gamma isoform
	IGF1R	insulin-like growth factor 1 receptor	ITPR2	inositol 1,4,5-triphosphate receptor, type 2
	AURKA	aurora kinase A	/	/
	SLK	STE20-like kinase (yeast)	/	/
p53 signalling pathway	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	CASP8	caspase 8, apoptosis-related cysteine peptidase
	CDK4	cyclin-dependent kinase 4	ATM	ataxia telangiectasia mutated (includes complementation groups A, C and D)
	GADD45A	growth arrest and DNA-damage-inducible, alpha	CDK2	cyclin-dependent kinase 2
	TNFRSF10B	tumor necrosis factor receptor superfamily, member 10b	CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
	SESN1	sestrin 1	TP53I3	tumor protein p53 inducible protein 3
Pathways in cancer	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	CASP8	caspase 8, apoptosis-related cysteine peptidase
	CDK4	cyclin-dependent kinase 4	CDK2	cyclin-dependent kinase 2
	ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
	CDH1	cadherin 1, type 1, E-cadherin (epithelial)	PDGFRA	platelet-derived growth factor receptor, alpha polypeptide
	RASSF1	Ras association (RalGDS/AF-6) domain family 1	STAT5B	signal transducer and activator of transcription 5B
	MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	HDAC1	histone deacetylase 1
	IGF1R	insulin-like growth factor 1 receptor	FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)

	HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A member 1	LAMA4	laminin, alpha 4
	HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1	DCC	deleted in colorectal carcinoma
	CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa	EPAS1	endothelial PAS domain protein 1
	FGF11	fibroblast growth factor 11	/	/
	TCEB1	transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C)	/	/
	LAMB1	laminin, beta 1	/	/

Table 5.9 KEGG pathways and corresponding genes disrupted in cSCC cell lines due to differential methylation or transcription. DifMET = differentially methylated genes, DifEXP=differentially expressed genes.

Because GO terms annotation can be extensive and too generic for functional comparison (e.g. "*membrane*"), I used only GO terms that fall in the category of "*biological processes*" and "*molecular function*" for the following analysis.

The number of significantly enriched GO terms in the list of differentially methylated and expressed genes is 92 and 99, respectively. Of those GO terms, 23 overlap.

The list of overlapping GO terms with corresponding number of annotated genes in each group and with overlapping genes is provided in Table 5.10, and a graphical overview is shown in Figure 5.24.

GO term	DifEXP genes	DifMET genes	Overlap
ANATOMICAL_STRUCTURE_DEVELOPMENT	47	26	TWIST1
ANATOMICAL_STRUCTURE_MORPHOGENESIS	16	9	TWIST1
BIOPOLYMER_METABOLIC_PROCESS	76	24	TWIST1, IVL
BIOSYNTHETIC_PROCESS	19	11	/
CALCIUM_ION_BINDING	8	6	GSN
CATABOLIC_PROCESS	13	7	/
CATION_BINDING	14	8	GSN
ESTABLISHMENT_OF_LOCALIZATION	36	22	AQP9
ION_BINDING	16	10	GSN
LIPID_METABOLIC_PROCESS	16	8	/
MULTICELLULAR_ORGANISMAL_DEVELOPMENT	41	28	TWIST1
NEGATIVE_REGULATION_OF_METABOLIC_PROCESS	13	7	GSN, TWIST1
ORGAN_DEVELOPMENT	29	19	TWIST1
POSITIVE_REGULATION_OF_BIOLOGICAL_PROCESS	33	17	/
POSITIVE_REGULATION_OF_CELLULAR_PROCESS	32	15	/
PROTEIN_METABOLIC_PROCESS	59	20	CCT6A, GSN, IVL
REGULATION_OF_BIOLOGICAL_QUALITY	19	14	GSN
REGULATION_OF_CELLULAR_METABOLIC_PROCESS	29	14	GSN, TWIST1
REGULATION_OF_METABOLIC_PROCESS	29	15	GSN, TWIST1
RESPONSE_TO_CHEMICAL_STIMULUS	16	12	AQP9
SIGNAL_TRANSDUCTION	46	46	/

SYSTEM_DEVELOPMENT	38	25	TWIST1
TRANSPORT	31	20	AQP9

Table 5.10 Overlapping GO terms between differentially methylated (DifMET) and differentially expressed (DifEXP) genes. Notably, only four genes overlapped between the two datasets: TWIST1, IVL, AQP9 and GSN.

Given that the number of genes in the differentially expressed group is much higher than in the differentially methylated group, the number of annotated genes is inevitably higher in this group.

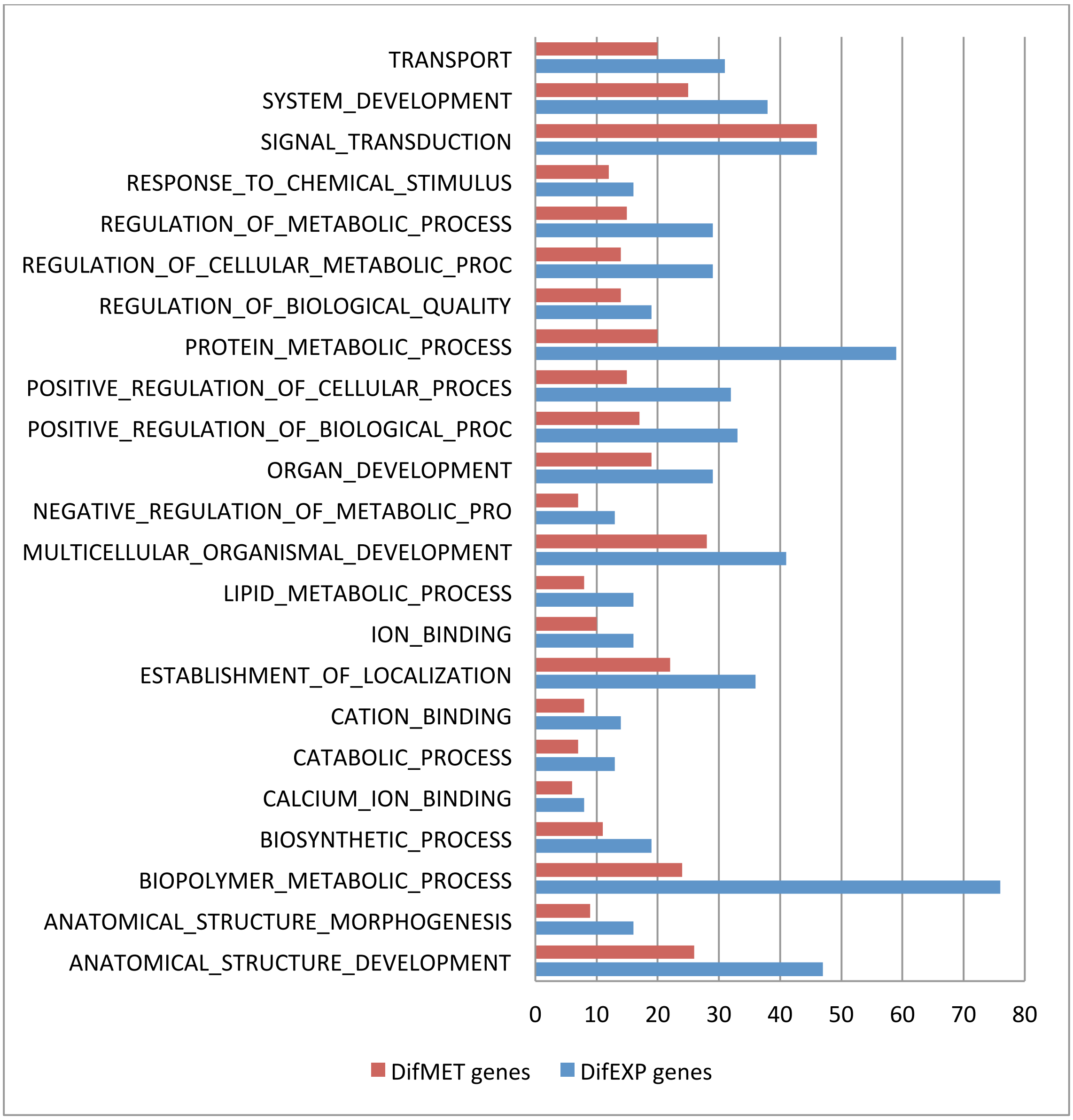


Figure 5.24 Graphical representation of the number of genes annotated with overlapping GO terms in gene expression and gene methylation data.

5.3.8.4. Comparison of transcriptional profile of RDEB-cSCC and non-RDEB cSCC cell lines

Given that previous methylation analysis has discovered a subset of of genes that is differentially methylated in RDEB-cSCC cell lines, I compared transcriptional profile of the two cSCC line types. Because the number of RDEB-cSCC cell lines is relatively low (3), I compared those to non-RDEB cSCC cell lines present on the same array chip in order to avoid a batch effect (6). The same cell lines were used for methylation profiling.

Despite of significant differences in methylation between RDEB and non-RDEB cSCC cell lines in several methylation loci, only 3 probes were found to be differentially expressed in RDEB-cSCC cell lines, and none of these represented an annotated gene.

5.3.8.5. Comparison of hESC, PHK and cSCC cell lines transcriptional profiles

Human embryonic stem cells represent the most undifferentiated cellular stage. A comparison with PHK has revealed 1092 probes to be differentially expressed in PHK compared with hESC on our gene expression array. As expected, among the most overexpressed genes in PHK were keratin 5, 6A, 6B, 14, 16, and keratin 17, and conversely, the most under-expressed genes included markers of pluripotency such as LIN28, POU5F1 (OCT4), SOX2, both molecules are critical for stem cell maintenance, and also podocalyxin that was previously shown to be highly expressed in hESC (Tateno et al., 2013).

Comparison between cSCC cell lines and hESC has revealed 1967 differentially expressed probes.

I applied a similar approach described in 5.2.4.1. to identify “stem cell signatures” in the transcriptional profile of cSCC cell lines: I looked for genes that were differentially expressed in cSCC cell lines compared to PHK, and in hESC compared to PHK, but not in hESC compared to cSCC cell lines. This approach identified 13 genes as a putative transcriptional hESC signature in cSCC cell lines. Only 2 of the genes are over-expressed in cSCC cell lines. The genes are listed in Table 5.11, and Figure 5.25 shows log2 tranformed expression levels in all samples.

Gene	Name	Expression status	Average log2 in PHK	Average log2 in hESC	Average log2 in cSCC lines
ABCA12	ATP-binding cassette, sub-family A (ABC1), member 12	Downregulated	7.756347901	5.781304471	6.247462963
ANKRD13A	ankyrin repeat domain 13A	Downregulated	10.20392824	8.330860529	9.032076911
ARL8B	ADP-ribosylation factor-like 8B	Downregulated	10.58582747	9.015457599	9.48552854
BTG2	BTG family, member 2	Downregulated	9.234047781	6.630964878	7.13495487
C9orf150	chromosome 9 open reading frame 150	Downregulated	8.004069303	5.876505306	6.273424674
CWH43	cell wall biogenesis 43 C-terminal homolog (S. cerevisiae)	Downregulated	9.744739637	5.553054785	6.528415436
HSPA4L	heat shock 70kDa protein 4-like	Downregulated	8.837667108	6.654665628	7.407126652

HSPC159	galectin-related protein	Downregulated	10.19507674	8.29014429	7.863605111
PLEKHA1	pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1	Downregulated	10.53528504	9.136564582	9.683045639
RPS26	ribosomal protein S26	Overexpressed	6.240519279	7.698226226	7.16242525
SLC20A2	solute carrier family 20 (phosphate transporter), member 2	Downregulated	9.676992731	7.705310222	8.278802779
TRAK1	trafficking protein, kinesin binding 1	Downregulated	9.989411899	7.494696762	8.153990433
TWIST1	twist homolog 1 (Drosophila)	Overexpressed	6.067105355	7.960261748	7.925830216

Table 5.11. hESC transcriptional signature in cSCC cell lines.

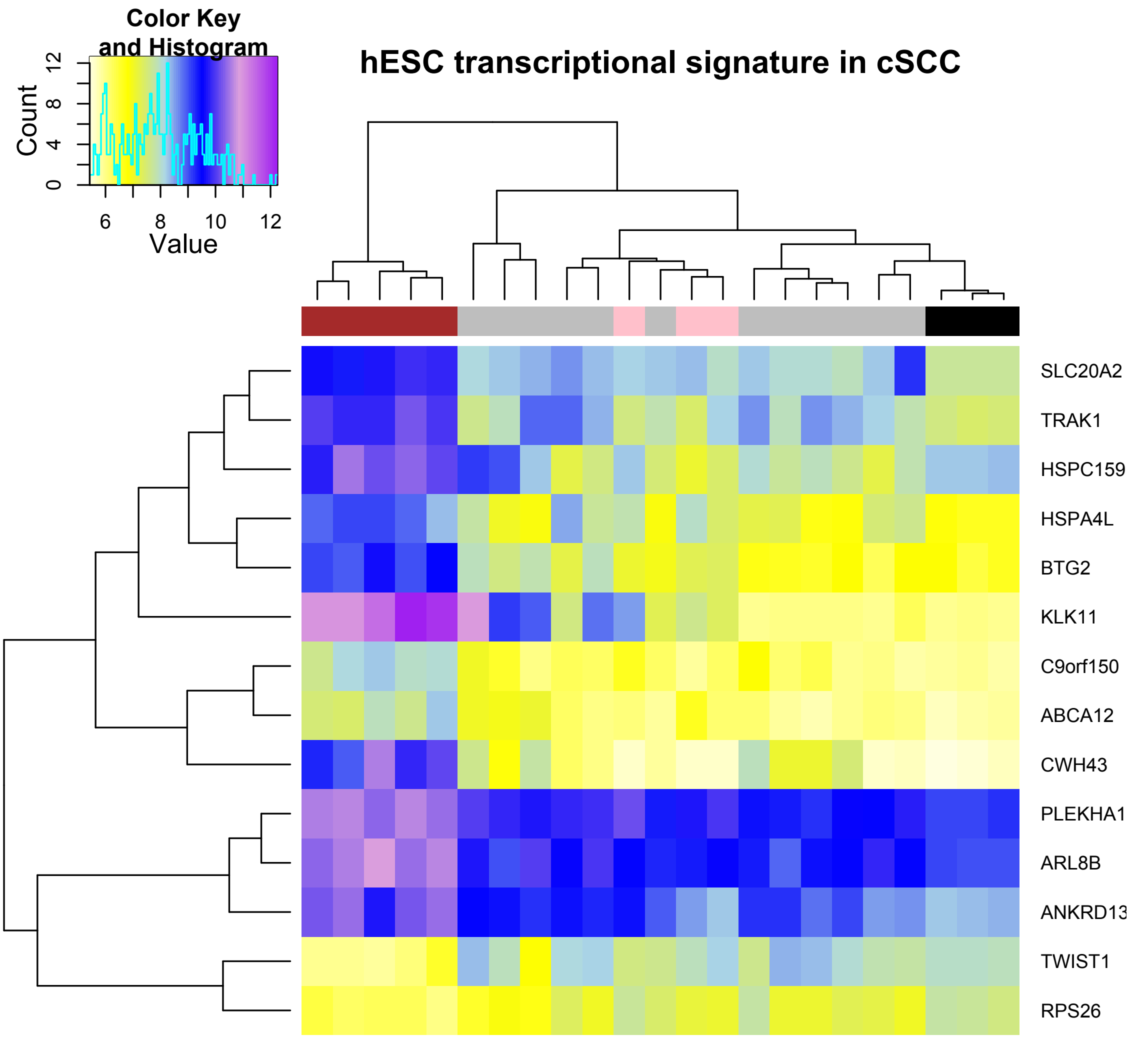


Figure 5.25 hESC transcriptional signature in cSCC cell lines. Individual rectangles depict log2 transformed expression values for given probes in each sample. PHK=red, hESC=black, RDEB cSCC cell lines=pink, non-REDEB cSCC cell lines=grey.

I also aimed to identify a unique set of genes that is exclusively over- and undertranscribed in cSCC cell lines in comparison to hESC and PHK. Such genes do not change their transcription level in PHK compared to hESC, but change transcription levels in cSCC cell lines in comparison with both hESC and PHK. This approach removes genes that are differentially expressed due to physiological differentiation, or non-malignant undifferentiated state, and leaves genes that represent a “cancer-specific” signature.

Only 29 genes fall within this category (listed in Table 5.12), of which 8 are overexpressed, and 21 are underexpressed in cSCC cell lines. Figure 5.26 shows log2 transformed expression levels for each probe in individual samples.

Symbol	Name	Expression status	Average log2 in PHK	Average log2 in hESC	Average log2 in cSCC lines
ACVR2A	activin A receptor, type IIA	Downregulated	7.174753382	7.199160402	6.302883028
ANKRD35	ankyrin repeat domain 35	Downregulated	7.762226648	7.74241036	5.981989873
COL22A1	collagen, type XXII, alpha 1	Downregulated	7.631004719	7.598820483	5.933940829
EPHB6	EPH receptor B6	Downregulated	6.523747094	6.558252646	5.626251992
EPHX3	epoxide hydrolase 3	Downregulated	7.761661576	8.312469864	5.976064913
FAM46B	family with sequence similarity 46, member B	Downregulated	9.935372259	10.79121751	7.379594766
ITM2A	integral membrane protein 2A	Downregulated	9.897791418	9.068175733	6.299472246
LPHN2	latrophilin 2	Downregulated	8.362233004	10.03058091	6.207604557
LZTS1	leucine zipper, putative tumor suppressor 1	Downregulated	6.207306169	6.67818831	5.720718317
NDN	necdin homolog (mouse)	Downregulated	8.015988227	9.808595573	6.15872661
NISCH	nischarin	Downregulated	8.678876908	9.408698212	7.407153546
NYNRIN	NYN domain and retroviral integrase containing	Downregulated	8.469562988	7.966256456	6.102175778
P4HTM	prolyl 4-hydroxylase, transmembrane (endoplasmic reticulum)	Downregulated	7.81567745	8.10233413	6.581043205
PAMR1	peptidase domain containing associated with muscle regeneration 1	Downregulated	7.481602661	7.325083649	5.682580992
PAPSS1	3'-phosphoadenosine 5'-phosphosulfate synthase 1	Downregulated	10.26674493	10.36927448	9.270910268
PPP2R2B	protein phosphatase 2 (formerly 2A), regulatory subunit B, beta isoform	Downregulated	8.714079329	10.6676766	6.38705536
PRICKLE1	prickle homolog 1 (Drosophila)	Downregulated	8.484905315	9.864984913	6.738059773
RPS27L	ribosomal protein S27-like	Downregulated	11.89427539	12.33474575	10.83371845
SESN1	sestrin 1	Downregulated	7.846310241	8.048683661	6.795726138
VWA5A	von Willebrand factor A domain containing 5A	Downregulated	8.448715664	7.395725704	6.455520204
ZNF334	zinc finger protein 334	Downregulated	6.752018712	7.310928817	5.954087262
C13orf27	chromosome 13 open reading frame 27	Overexpresssed	8.144870542	7.938208813	9.291159212
C1orf59	chromosome 1 open reading frame 59	Overexpresssed	6.788040774	7.233437861	9.350337401
C6orf150	chromosome 6 open reading frame 150	Overexpresssed	5.687870702	5.561222239	7.857132716
HOXB7	homeobox B7	Overexpresssed	5.924168288	5.672006308	8.73611203
NME2	non-metastatic cells 2, protein (NM23B) expressed in	Overexpresssed	7.445045761	7.061602502	8.365042574
RILPL2	Rab interacting lysosomal protein-like 2	Overexpresssed	6.984356728	6.418957327	8.098413922

UQCRB	ubiquinol-cytochrome c reductase binding protein	Overexpresssed	9.502009046	9.121772702	10.59440563
WDR66	WD repeat domain 66	Overexpresssed	7.46809362	5.860627283	10.14745016

Table 5.12 List of genes that represent cancer-specific transcriptional signature in cSCC cell lines.

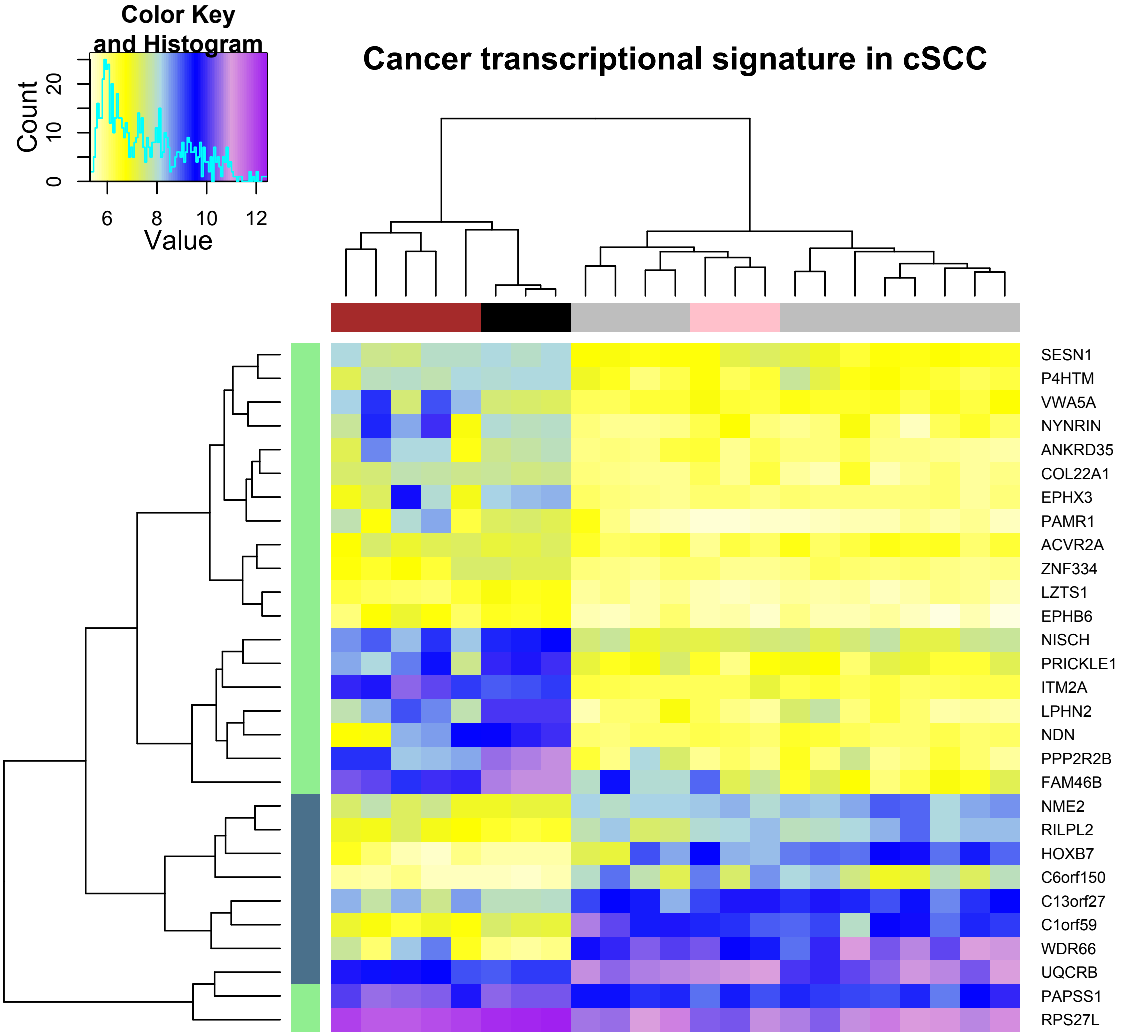


Figure 5.26 Log2-transformed transcriptional levels of genes that are differentially expressed in cSCC cell lines compared to both hESC and PHK, but not in PHK compared to hESC. Top cluster: PHK=red, hESC=black, RDEB cSCC cell lines=pink, non-REDEB cSCC cell lines=grey. Left cluster: skyblue=probes that are overexpressed in cSCC cell lines, light green=underexpressed probes in cSCC cell lines.

5.4. Discussion

I have performed the first genome-wide study of DNA methylation and transcriptional profiling of primary cutaneous SCC cell lines generated from non-RDEB and RDEB-cSCC samples using a microarray high-density oligonucleotide chips. I compared the methylation and transcriptional profile to primary cultured keratinocytes and hESC to test the hypothesis that cSCC cell lines gain stem-cell like properties that drive their malignant phenotype.

5.4.1. 371 genes are differentially methylated between normal primary keratinocytes and cSCC cell lines and include homeobox genes

The comparison of normal primary keratinocytes with cSCC cell lines lead to the identification of 371 differentially methylated genes. Of these genes, one was previously described to be epigenetically silenced in cutaneous SCC (p16, CDKN2A) (Brown et al., 2004). Putative tumour suppressor TP53 promoter was also previously shown to be hypermethylated in cSCC, and I detected CpG hypermethylation of tumour protein p53 inducible protein 3, TP53-contolled gene, which indicates that TP53 regulatory network may be disrupted at multiple levels through various molecules. Moreover, I discovered several oncogenes and tumour suppressor genes not previously described to be involved in cSCC to be differentially methylated in our samples, including MRAS, SERPINA3, SIM1, SPARCL1, REGL, and CCR3. We also discovered differential methylation of genes that have not been previously implicated in cancer, but that may play a role in cSCC development including SNRK, which is regulated by tumour suppressor 14-3-3 via STK11, and was hypermethylated in our tumour samples, so was TANK, a NFkappaB activator. DYSF probe, which is upregulated in sarcolemma (the cell membrane of cells that form a striated muscle fibers) wound healing, was also hypomethylated, and so was LMLN, a metallopeptidase that promotes cellular invasion. Hypomethylation of TWIST1 is a novel finding in cSCC, and suggests that EMT may play a prominent role in cSCC carcinogenesis. These genes represent potential novel target genes in skin cancer research.

Four homeobox genes were also differentially methylated in comparison between skin and tumours including PAX4, which has previously been implicated in cancer development. Methylation of homeobox genes has been previously shown in breast cancer (Tommasi et al., 2009), testicular cancer (Lind et al., 2006), and lung cancer (Rauch et al., 2007). In addition to aberrant methylation, deregulated expression of homeobox genes in cancer has been reported by many studies (Abate-Shen, 2002), and our findings warrant further functional evaluation of homeobox genes in cSCC.

5.4.2. Genome-wide methylation profiling of cSCC cell lines distinguishes between cSCC cell lines derived from well-differentiated tumours and less differentiated cSCCs

Clustering based on correlation of methylation profile of 9 cSCC cell lines was able to segregate RDEB-cSCC cell lines as a distinct group, and segregate additional cSCC cell lines based on original tumour histology. This finding was consistent with previous genotyping of cSCC; however, transcriptional profiling of cSCC cell lines did not detect such trend. This may be explicable by the general lack of correlation between epigenetic and transcriptional differences, or – since in the previous chapter, such trend was indeed detected by gene expression profiling of cSCC clinical samples - by transcriptional “homogenisation” of cSCC cell lines due to cell culture, and a greater impact of technical processing on cultured cells, such as treatment with trypsin and centrifugation prior to RNA collection in comparison with tumour collected from patients. Additionally, both observations may be explicable by low number of samples, and are subject to further validation.

5.4.3. EB cell lines have a distinct methylation profile to non-EB lines but not distinct transcriptional profile

We have shown that DNA methylation pattern is crucially different in tumour cell lines derived from EB patients in comparison with non-EB patients and that several developmental and cancer-related genes are significantly differentially methylated between these neoplastic entities. This difference in methylation profile may at least partially corroborate distinct clinical behaviour of EB-associated SCC. Namely the epigenetic silencing of potential tumour suppressor DOK1, GATA4, invasion suppressor CRMP1 and MFRP hypomethylation, a gene implied in Wnt pathway signalling, may be accountable for more aggressive propensities of EB SCC. KEGG Pathway analysis pointed to differential methylation of Gap junction and several metabolic pathways, and GO term analysis further stressed the involvement of metabolism-related genes. GO term analysis also pointed to significant overrepresentation of genes involved in neural processes, DNA repair and ERK pathway in hypermethylated group of genes.

Unexpectedly, transcriptional profiling revealed no differentially transcribed genes between the two kinds of cSCC cell lines. This may be explained by methylation differences representing a mere array artefact due to higher passage number of RDEB-cSCC cell lines, or due to their genetics, rather than epigenetics, or due to biased statistical inference caused by including only three RDEB-cSCC cell lines in the dataset. Further validation of differences between RDEB and non-RDEB cSCC on both epigenetic and transcriptional level is required.

5.4.4. Validation of 27K methylation array with bisulfite sequencing confirms direction of methylation change in 46.7% of tested genes

Bisulfite sequencing data of laser-capture microdissected samples were used to validate genes identified as differentially methylated in cSCC cell lines. 163 genes (46.7%) showed consistent direction of methylation change. Of 163 genes for which the data was validated, 81 of the validated genes were hypo- and 82 were hypermethylated. This finding suggests that there may be a statistical bias in the methylation data due to only three normal human keratinocytes samples used as control, in addition to possible cell culture array artifacts. However, this also indicates that many genes detected as differentially methylated in cSCC cell lines represent valid targets that correspond with clinical samples. Such finding is important namely for functional evaluation of methylation targets in the future.

5.4.5. Methylation and transcriptional profiling of cSCC cell lines shows little overlap in dysregulated genes, but a notable overlap in dysregulated pathways and GO terms

As expected, there was minimal overlap between dysregulated genes on methylation and transcriptional level (6.8% of differentially methylated genes, 3.8% of differentially expressed genes). Integration of transcriptional data with methylation based on ranges of methylation difference showed expected decrease in gene expression with increasing gene methylation in primary human keratinocytes, but expression levels of genes with the highest promoter methylation paradoxically increased. This indicates a more complex role of promoter methylation, in addition to transcription repression.

Pathways dysregulated both differentially expressed and methylated genes included Cell cycle, Oocyte meiosis, p53 signalling pathway and Pathways in cancer. Given the minimal overlap of genes between the two datasets, this observation suggests that while there may be a common regulatory disruption between the transcriptional and epigenetic control mechanisms, it is likely conveyed by different genes; possibly in a synergistic manner.

5.4.6. Comparison of methylation profile of primary normal human keratinocytes and cSCC cell lines with additional tissues shows concordance of primary normal human keratinocytes with normal skin and complete segregation of cSCC cell lines from other cultured cell types

An important finding is the high correlation of primary normal human keratinocytes (PHK) and normal skin methylation profile observed in our data. This observation indicates that primary keratinocytes cultured for a short period of time largely maintain the methylation profile of uncultured cells, and validates this model for further exploration of epigenetic regulation in normal skin.

Based on the correlation of the transcriptome of less differentiated cSCC samples and mesenchymal stem cells, I hypothesised that cSCC cell lines may show methylome correlation with less differentiated tissue types, and assessed this by a comparison with a series of additional cultured samples. While PHK showed correlation a degree of correlation with fetal lung fibroblasts (possibly due to the presence of fibroblasts in the keratinocyte cultures), cSCC cell lines formed an entirely independent cluster in this analysis, suggesting that their methylome is strictly cancer-specific. In addition to representing a valid biological finding, this observation may be due to technical aspects (batch effect) and due to the array bias, since it evaluates methylation of only a limited number of genomic regions, and also due to the low number of cSCC samples.

5.4.7. Detection of hESC signature in methylation and transcriptional data

Based on the observation of shared cellular properties between stem cells and cancer, such as dedifferentiation and migration propensity, I hypothesised that a certain subset of genes may functionally underpin this observation. To detect these genes, I identified genes that are differentially methylated/expressed in cSCC cell lines compared to normal skin, but are not differentially methylated/expressed in cSCC cell lines compared to hESC. Using this approach, I detected a surprisingly low number of genes corresponding to these criteria on both methylation and transcriptional level. At methylation level, one of the signature genes, ACTA2, a maker of differentiation and a p53 target, was hypermethylated in both cSCC cell lines and hESC, warranting further exploration of its role of dedifferentiation in cancer-specific context. At transcriptional level, TWIST1, an EMT driver, was upregulated in both hESC and cSCC cell lines, which indicates it may be a direct driver of the less differentiated phenotype of cSCC.

5.4.8. Conclustions and study limitations

Overall, our results provide the first genome-wide insight into cSCC methylation profile and provides clues with respect to possible molecular mechanisms involved in cSCC oncogenesis. We identified several potential targets for further investigation in skin cancer research. Moreover, we demonstrated that EB-derived cSCC are an independent subset of cSCC from epigenetics perspective and if further studies confirm this hypothesis, EB-derived cSCC may be approached as an independent biological entity with specific treatment options. The fact that the methylation status of genes in our set was not always concordant with previous results regarding its oncogenic or tumour-suppressor character indicates that methylation status of individual genes may often be tumour tissue-specific, and in addition to this, methylation has been shown to vary across different specimen of the same tumour type. If our findings are validated by further studies, these would point to considerations of individual tumour profiling and potentially individually-tailored medicine for cSCC patients in the future.

Limitations of our study include the fact that methylation was detected in cell culture samples, rather than primary tumours. Whether tissue culture impacts methylation profiling of samples is a question addressed in our following experiment (see below).

5.5. Future directions

Future steps include validation of differential methylation loci in cSCC cell lines using Illumina 450K array and an additional set of cell lines for biological validation. Additionally, proposed hESC signature genes need to be validated with additional techniques and datasets, including Illumina 450K and bisulfite sequencing data of cSCC clinical specimen described in the following chapters. Finally, if this signature is confirmed, functional analysis of these genes would provide insight into the phenotypical characteristics of cSCC tumours.

6. Identification of differentially methylated genes in a sequential series of normal skin, premalignant actinic keratoses and malignant cSCC clinical samples

6.1. Introduction

As discussed in the previous chapter, profound differences in methylation exist between primary human keratinocytes and primary cutaneous cSCC cell lines. To the author's knowledge, no comprehensive, genome-wide methylation study has been previously conducted on squamous cell carcinomas of the skin or its precursor lesion actinic keratosis, and candidate gene methylation studies are limited. Thus far, silencing of CDKN2A locus has been shown to be commonly inactivated due to promoter methylation (Brown et al., 2004, Murao et al., 2006) in cSCC. Other frequently methylated genes in cSCC include FOXE1 (Venza et al., 2010, Chiles et al., 2003b) and E-cadherin (Chiles et al., 2003a). On the other hand, hypermethylation of DAPK promoter hs been demonstrated not to be involved in cSCC development (Tyler et al., 2003).

Methylation profiling studies of other epithelial SCCs are also relatively limited. In head and neck SCC, MGMT, DAPK, RARbeta, MLH1, RASSF5 and MST1 genes are hypermethylated in more than 50% of samples compared to healthy tissue (Steinmann et al., 2009), and hypermethylation of CRABP2 has been linked to poorer disease-free survival (Calmon et al., 2009). In lung SCC, CpG island hypermethylation has revealed 162 sites of differential methylation (Park et al., 2005). The difference in methylation profile in Caucasian males suffering from lung SCC seems to be accountable for their shorter survival (Piyathilake et al., 2003). Additionally, methylation profiling of non-small cell lung cancer (NSCLC) using the 450K methylation array platform and integration of the data with gene transcription scores has revealed phenotypically distinct molecular tumour subtypes (Walter et al., 2012).

Methylation profiling of oral SCC revealed differential methylation of p16, cytoglobin and cyclin A1 and hypermethylation of CKMT1 (Onda et al., 2006). HPV-infected cervical SCCs have been associated with hypermethylation of DAPK1, MGMT, CADM1 and CDH13 (Henken et al., 2007) and higher methylation of cyclin A1 corresponds to higher tumour grade (Kitkumthorn et al., 2006) of this neoplasia. All these data indicate that changes in methylation patterns are associated with SCC development, and that concrete alterations are most likely site-specific. Additionally, no data on methylation patterns in actinic keratosis exist.

The aim of this study was to explore methylation differences in a series of clinical samples representing progression from normal skin to actinic keratosis and cSCC. Additionally, I aimed to validate differences in methylation between normal skin and cSCC using whole genome bisulfite sequencing of paired, laser-capture microdissected samples collected from male organ transplant recipients.

6.2. Materials and Methods

6.2.1. Clinical samples hybridised to the DNA methylation array

A total of 60 clinical samples were included in the study. The series consisted of 10 non-sun exposed and 10 sun-exposed skin samples, 20 actinic keratosis (AK) samples and 20 cSCC samples. The series was derived from 34 individual patients (details are provided in Table 6.1).

Tissue	Histology	Immune Status	Gender	Patient ID
cSCC	WD	IC	M	M1
AK		IC	M	M1
AK		IC	M	M2
NSE		IC	M	M2
SE		IC	M	M2
cSCC	MD	OTR	M	M3
cSCC	WMD	OTR	M	M4
AK		OTR	M	M4
AK		OTR	F	MF5
SE		OTR	F	MF5
AK		OTR	M	M6
cSCC	MD	CLL	F	MF7
AK		IC	M	M8
SE		IC	M	M8
cSCC	MD	OTR	M	M9
AK		OTR	M	M9
NSE		OTR	M	M9
cSCC	MD	OTR	M	M10
NSE		OTR	M	M10
cSCC	SCC	OTR	M	M10
AK		CLL	M	M11
NSE		CLL	M	M11
AK		IC	M	M12
NSE		IC	M	M12
SE		IC	M	M12
cSCC	WD	OTR	M	M15
AK		OTR	M	M15
NSE		OTR	M	M15
SE		OTR	M	M15
cSCC	WD	OTR	M	M16
cSCC	WD	OTR	F	MF17
cSCC	WD	IC	M	M18

cSCC	WD	IC	M	M19
AK		IC	M	M19
NSE		IC	M	M19
AK		IC	M	M20
SE		IC	M	M20
NSE		OTR	F	MF21
cSCC	WD	OTR	M	M22
cSCC	PD	OTR	M	M23
cSCC	WD	OTR	F	MF24
AK		OTR	F	MF24
SE		OTR	F	MF24
cSCC	WD	OTR	M	M25
AK		OTR	M	M26
NSE		OTR	M	M26
AK		OTR	M	M27
cSCC	WD	OTR	F	MF28
AK		OTR	F	MF28
AK		OTR	M	M29
AK		IC	M	M30
SE		IC	M	M30
cSCC	SCC	OTR	M	M31
cSCC	MD	IC	M	M32
SE		OTR	F	MF33
AK		OTR	F	MF33
cSCC	PD	OTR	M	M34
AK		IC	M	M34
NSE		IC	M	M34
SE		IC	M	M34

Table 6.1. Characteristics of 60 clinical samples used to detect methylation differences in a series of clinical specimen hybridised to the Illumina 450K methylation array. The histological subtype of 2 cSCCs was unknown.

6.2.2. Infinium HumanMethylation450 BeadChip

DNA processing and hybridisation was carried out in collaboration with Barts and the London Genome Centre. DNA isolated from undissected clinical specimen was bisulfite-converted using EZ DNA Methylation-Gold™ Kit (Zymo Research, United States), and hybridised to Illumina 450K DNA methylation array. The amount of total input DNA ranged from 40 to 550 ng of DNA, and a total of 14 ul of bisulfite-converted DNA was then amplified using random hexamer priming and Phi29 DNA polymerase followed by end-point fragmentation, and hybridised to the array at 48°C for 17 h, followed by single nucleotide extension according to the manufacturer's protocol. The beadchips were then washed and read using the Illumina iScan system (Illumina, United Stated), and the scanned raw intensities were stored in *.idat files.

Infinium HumanMethylation450 BeadChip is a state-of-art methylation array containing 485,764 probes that correspond to cytosine sites across the human genome. Probes are 50 bp-long and are linked with an address code of 23 bp in length that provides information of a bead location on the microarray chip. For Type I probes, the beads (one for the methylated and one for the unmethylated locus state) are bound to sequences complementary to either fully methylated or unmethylated genes (type I probes used for the older Human Methylation 27k BeadChip assay are described in greater in the previous chapter), while Type II probes are bound with a single bead type and the methylated or unmethylated state is determined during the single base extension following DNA hybridisation.

The vast majority of these cytosine sites (99.3%) represent CpG dinucleotides, and the remaining sites (0.7%) are CNG targets, where "N" is any nucleotide. Additionally, 30.9% of CpGs are located in CpG islands, 23% in CpG shores (sequences distant from CpGs up to 2 kb), 9.7% in CpG shelves (sequences located from 2-4 kb from CpGs) (Irizarry et al., 2009) and "Open Sea" areas which represent isolated CpGs in the genome (Sandoval et al., 2011). Figure 6.1 depicts positional relationships of these elements.

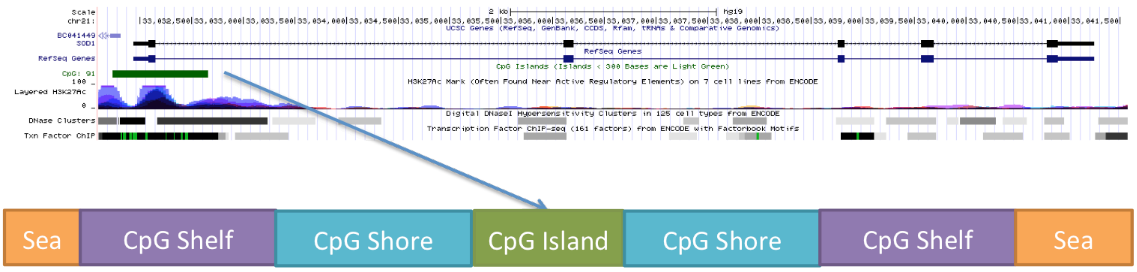


Figure 6.1 CpG Island and the positional relationship of CpG shore (blue), CpG Shelf (purple) and so-called "Open Sea" areas (orange) to CpG islands. CpG islands are defined as sequences more than 200 bp in length with CG representing more than half the sequence composition. CpG Shores are located within 2 kb distance from a CpG island, while CpG shelves are located 2-4 kb away from a CpG. Open Sea CpGs are isolated CpG areas in the genome.

6.2.3. Methylation array quality control

Quality control of raw data was carried out in collaboration with Dr Thomas Down[16], the remainder of the analyses was carried out by the author. DNA Methylation array quality control (QC) was carried out with the Minfi package (Aryee et al., 2014). The raw data values contained withtin idat files and a sample sheet containing information about the experimental layout were read into R

---

[16] The Gurdon Institute, University of Cambridge, United Kingdom.

and several QC plots using control probes present on the array were generated to determine signal quality of each array prior to normalisation (Figure 6.2). Additionally, density plots of individual arrays were plotted to observe bimodal distribution of probes (Figure 6.3). A total of three samples out of 60 (5%) were excluded from further analysis based on the QC: one AK sample (leaving 19 AK samples for further analysis) and 2 tumour samples (leaving 18 tumour samples for subsequent analysis). All normal skin samples (a total of 20) passed the quality control.

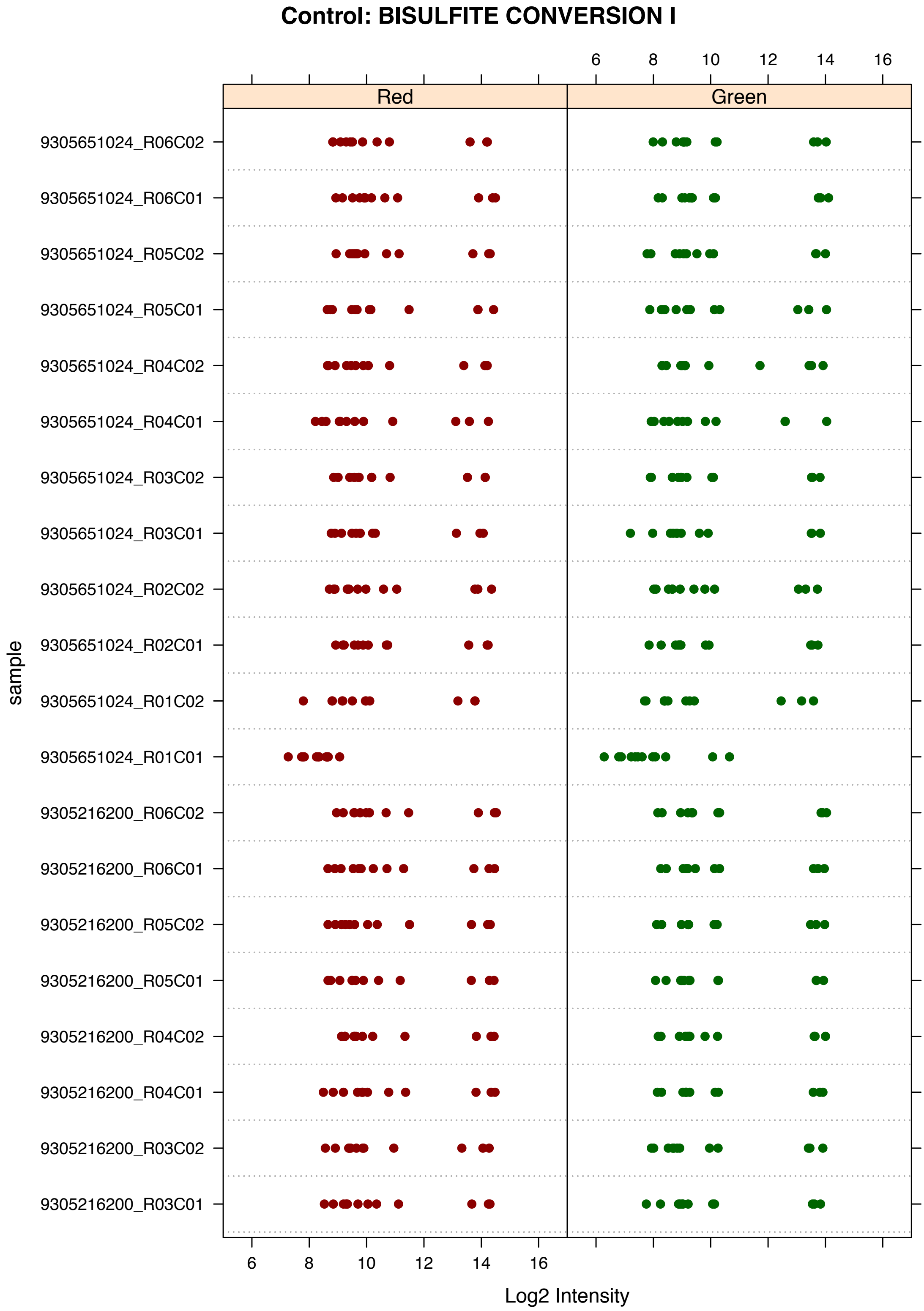


Figure 6.2 Log2 intensity of bisulfite conversion control probes for either green or red channel present on the 450K methylation array. Each sample is evaluated in two different colour channels (red or green) and the array contains probes that measure the efficacy of bisulfite conversion (each dot in the plot corresponds to a control probe). This plot shows that sample R01C01 is poorly bisulfite-converted, while the remaining samples pass this control step.

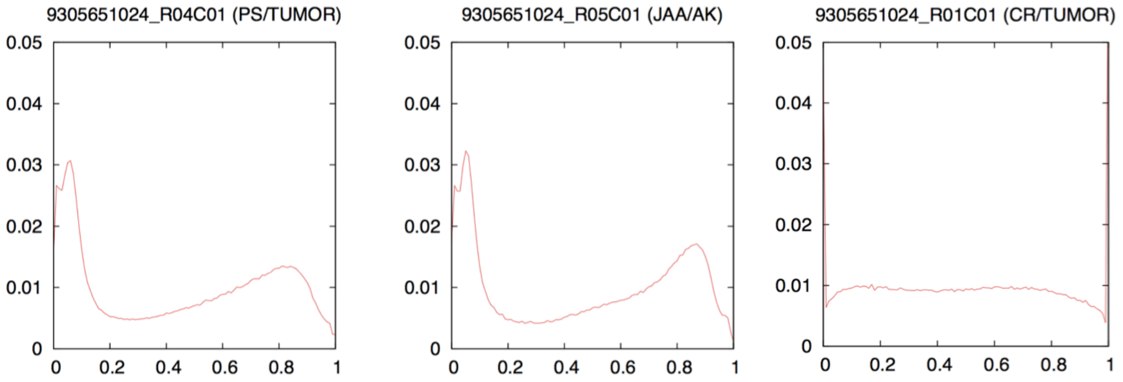


Figure 6.3 Density plot showing bimodal distribution of beta values on the x axis and the density on the y axis in a tumour sample (left), AK sample (middle) of proper shape in both plots that passes the quality control. A poor quality plot of a tumour sample is shown on the right. This sample was excluded from further analysis.

6.2.4. Removal of probes associated with SNP, raw data normalisation and determination of differentially methylated sites

The array is known to contain probes that are associated with SNP (Yousefi et al., 2013) which are likely to affect DNA methylation measurement, and these were filtered out prior to normalisation. Additionally, given the mixture of male and female participants in the study, probes corresponding to loci on the X and Y chromosome were removed. The raw data wereas then normalised using the quantile normalisation function in the *preprocessCore* R package and the differentially methylated regions were called using Welch's t-test using the *genefilter* package. P value was then adjusted using the Benjamini-Hochberg method in the *stats* package. Density plot of all samples pre- and post-quantile normalisation were generated with the *affy* package and are shown in Figure 6.4 and Figure 6.5, respectively.

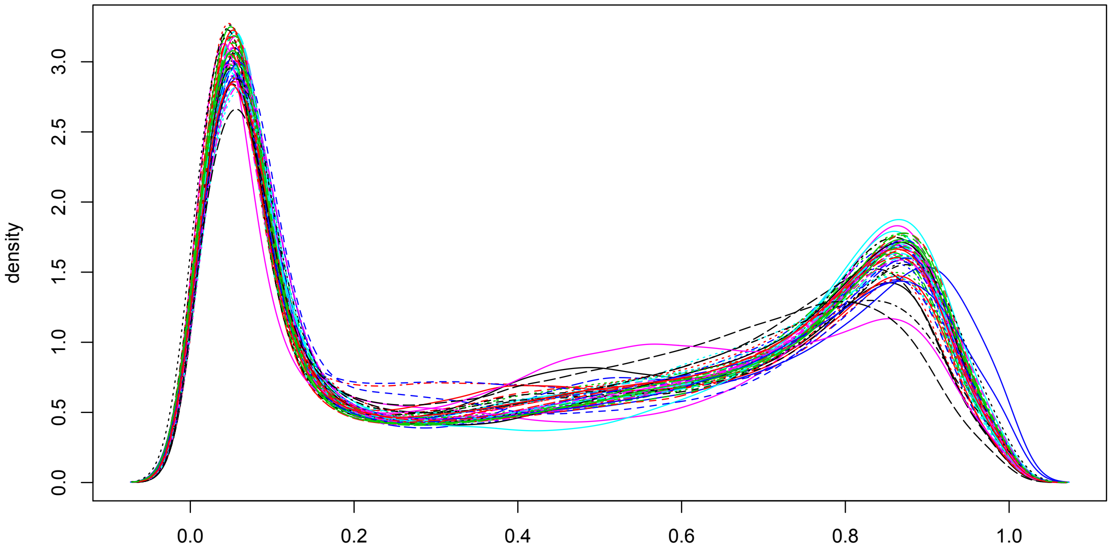


Figure 6.4 Density plot of all samples that passed QC prior to normalisation. All samples show proper bimodal distribution of p values, yet a certain level of noise is apparent, whereby the lines do not overlap perfectly. SNP probes are filtered in these data. X axis=beta values.

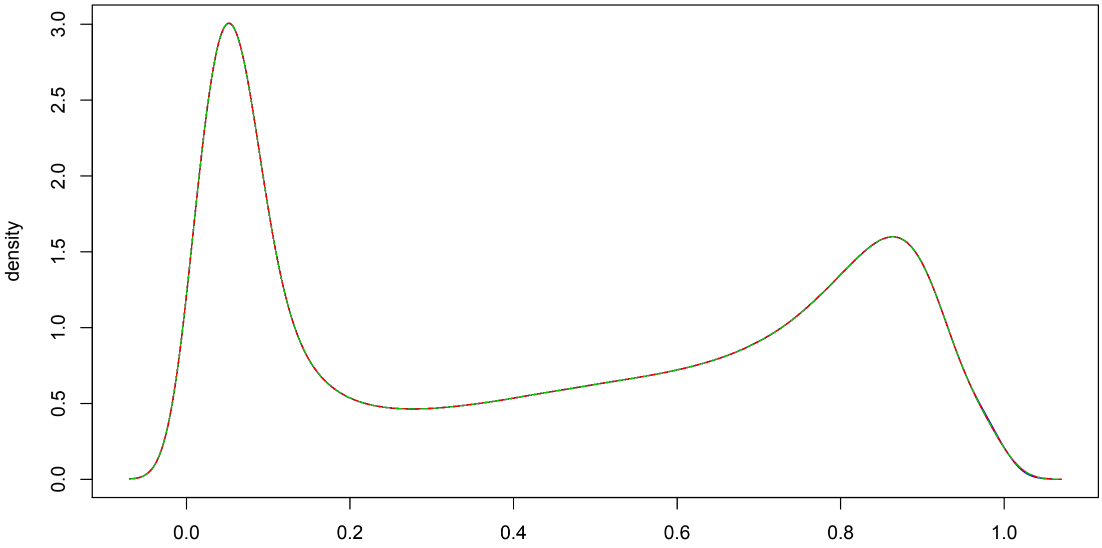


Figure 6.5 Density plot of all samples post-normalisation shows perfect overlap of beta values across all samples. Bimodal distribution of beta-values has been maintained post-normalisation.

6.2.5. Positional gene signature

Positional gene signature was obtained with Molecular Signature Database (MSigDB) of the Broad Institute of Massachussetts Institute of Technology (http://www.broadinstitute.org/gsea/msigdb). The database uses an overlap analysis between a gene set provided by the user and genes reported to be in cytobands of chromosomes based on HUGO, October 2006, and Unigene, build 197.

6.2.6. Integration with expression data

Normalised log2-transformed signal values of expression probles corresponding to differentially methylated genes on the Human Genome U133A 2.0 Affymetrix Array were extracted and mean log2 ratio calculated to observe correlation between differential methylation status and expression changes.

6.2.7. Correlation with immune status and cSCC histological typing

Association of methylation values with the immune status of patients and with cSCC histological subtype was carried out using the package *CpGassoc*. This package measures association between a phenotype of interest (immune status, histological subtyping).

6.2.8. Transcription Factor Targets Analysis

Enrichment for motifs among hyper and hypomethylated promoters was obtained with Molecular Signature Database (MSigDB) of the Broad Institute of Massachussetts Institute of Technology (http://www.broadinstitute.org/gsea/msigdb). The database uses curated TF binding sites defined in TRANSFAC (ver 7.4) to detect genes that share a TF binding site.

6.3. Results

6.3.1. Hierarchical clustering of normalised beta values separates samples based on histology

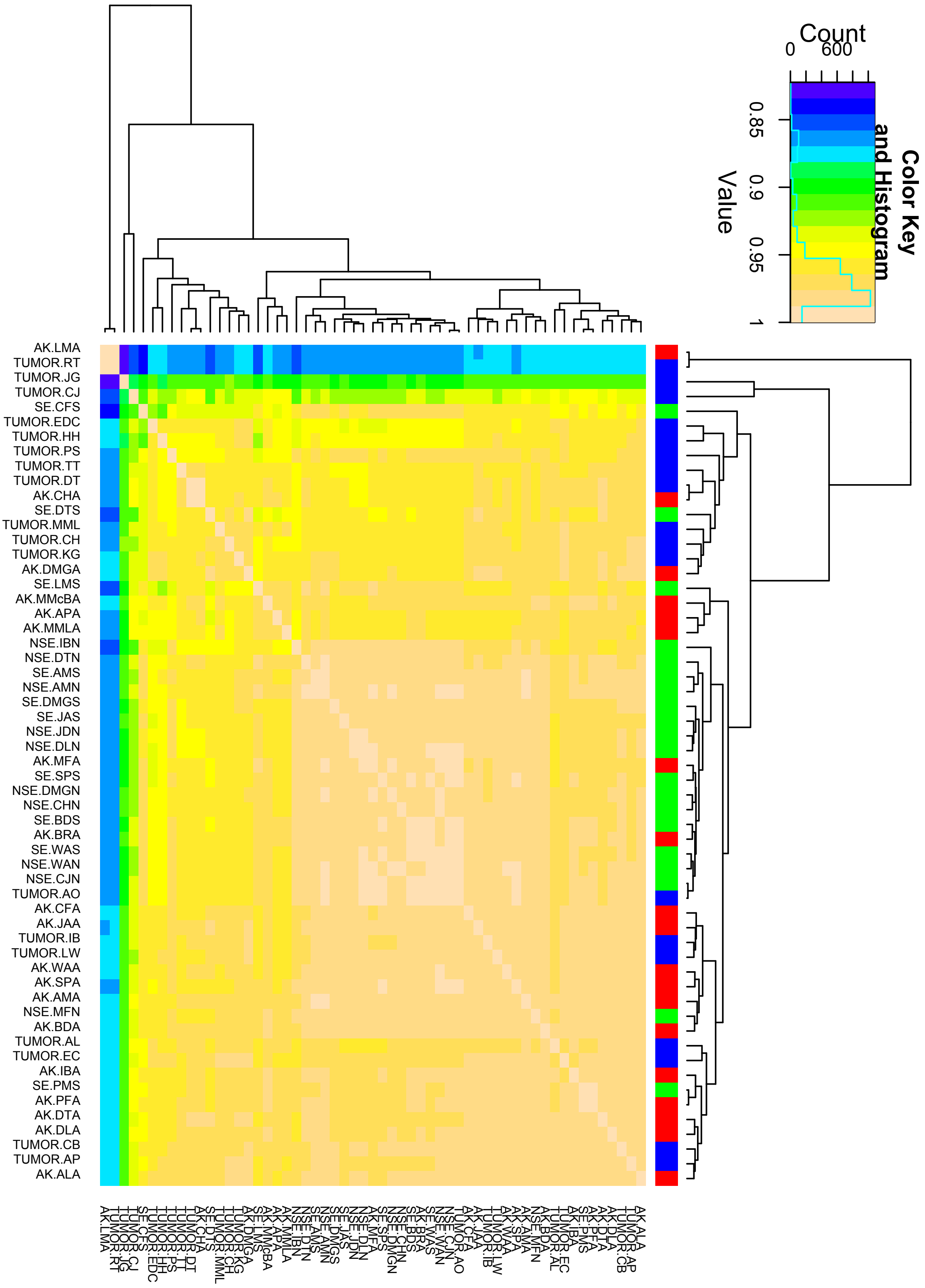


Figure 6.6 Hierarchical clustering of genome-wide methylation profile of all samples based on Pearson's correlation. Normal skin samples (green) form a distinct central cluster, 11 cSCC (blue) samples form a cluster on the top, and 11 AK(red) form a mixed cluster on the bottom.

As shown in Figure 6.6, Pearson's correlation of samples leads to distinct clustering of 15/20 skin samples in the central cluster, 11/18 cSCC samples in the top cluster and 11/19 AK samples in the bottom cluster. Although the distinction of samples based on their beta-values is not absolute, it is sufficient to expect differentially methylated genes and conduct group-wise comparisons.

6.3.2. Sun-exposed and non-sunexposed skin are not differentially methylated

I hypothesised that there may be minimal if any differences in genome-wide methylation of sun-exposed (SE) and non-sun exposed skin (NSE). Yet as shown in Figure 6.7, three sun-exposed skin samples formed a cluster on the left. Thus I first compared the methylation profile of 10 NSE and 10 SE skin samples. This comparison revealed no statistically significant differences in methylation across these two tissue categories. Thereby all further detections of differentially methylated genes in group-wise comparisons were conducted with combined NSE and SE data representing normal skin (20 samples), which increased the power of subsequent comparisons.

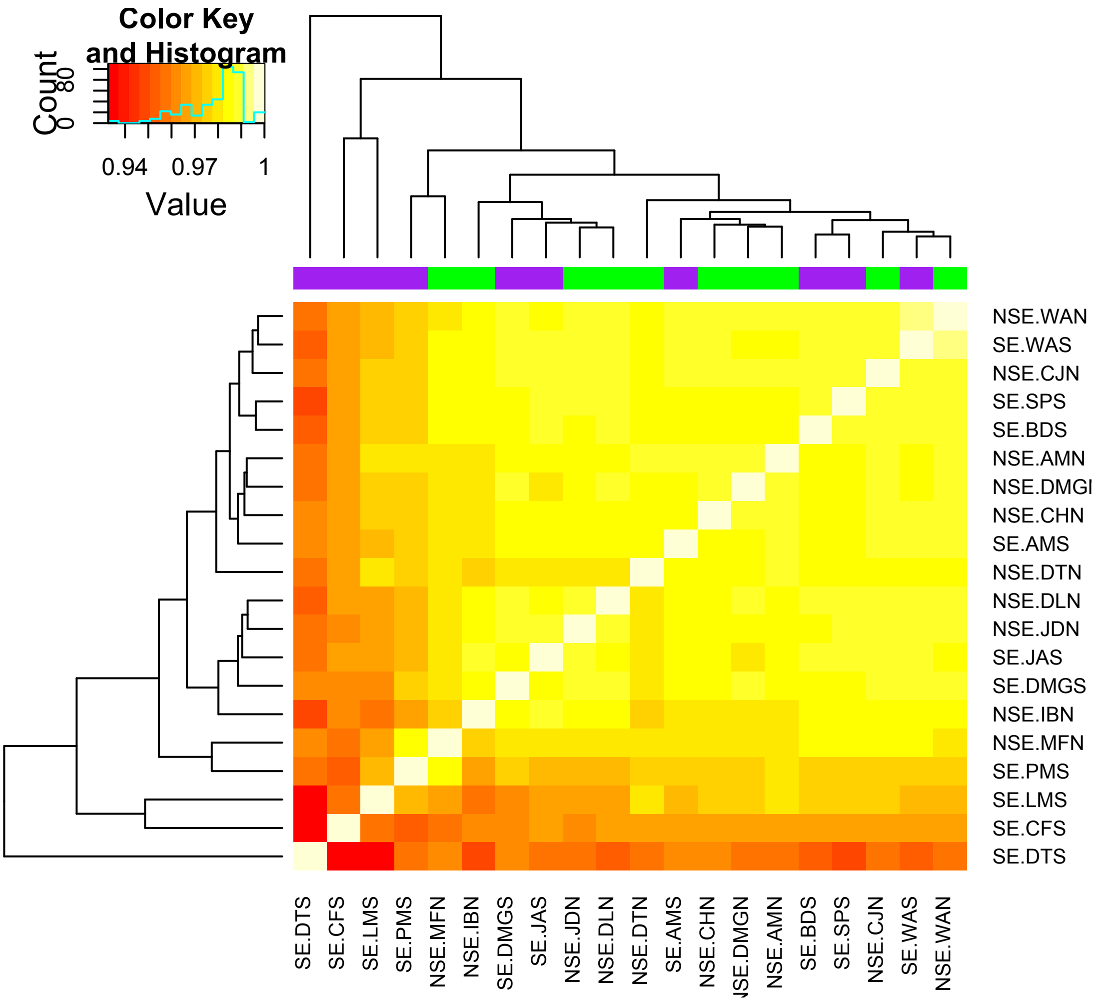


Figure 6.7 Hierarchical clustering of sun-exposed (purple) and non-sun exposed (green) skin samples based on quantile-normalised beta-values. This graph shows that the sun-exposed and non-sunexposed samples are largely interspersed, and show high correlation of their methylation profile. The top-right cluster is formed by paired samples from the same patient (WA), and similarly the second top-right cluster contains 2 paired samples (AM). The correlation model is not aware of patient ID.

6.3.3. Differentially methylated genes in group-wise comparisons

The following comparisons were carried out: skin vs. AK, skin vs. cSCC, AK vs. cSCC. The number of differentially methylated probes (DMP) detected in each comparison after p value adjustment is shown in Table 6.2, along with coefficient of determination for mean beta values that describes how well the two variables fit the regression model.

Cut-off	DMP Skin-AK	DMP Skin-cSCC	DMP AK-cSCC
P val <0.0001	2	5137	0
P val <0.01	8697	35175	0
P val <0.05	36497	68928	148

Coefficient of determination (r^2)	0.9882928	0.9762108	0.9922041

Table 6.2 Number of differentially methylated probes (DMP) detected in each comparison with a different adjusted p value cut-off.

This largely reflects the high correlation of mean methylation values in AK and cSCC samples shown in Figure 6.8.

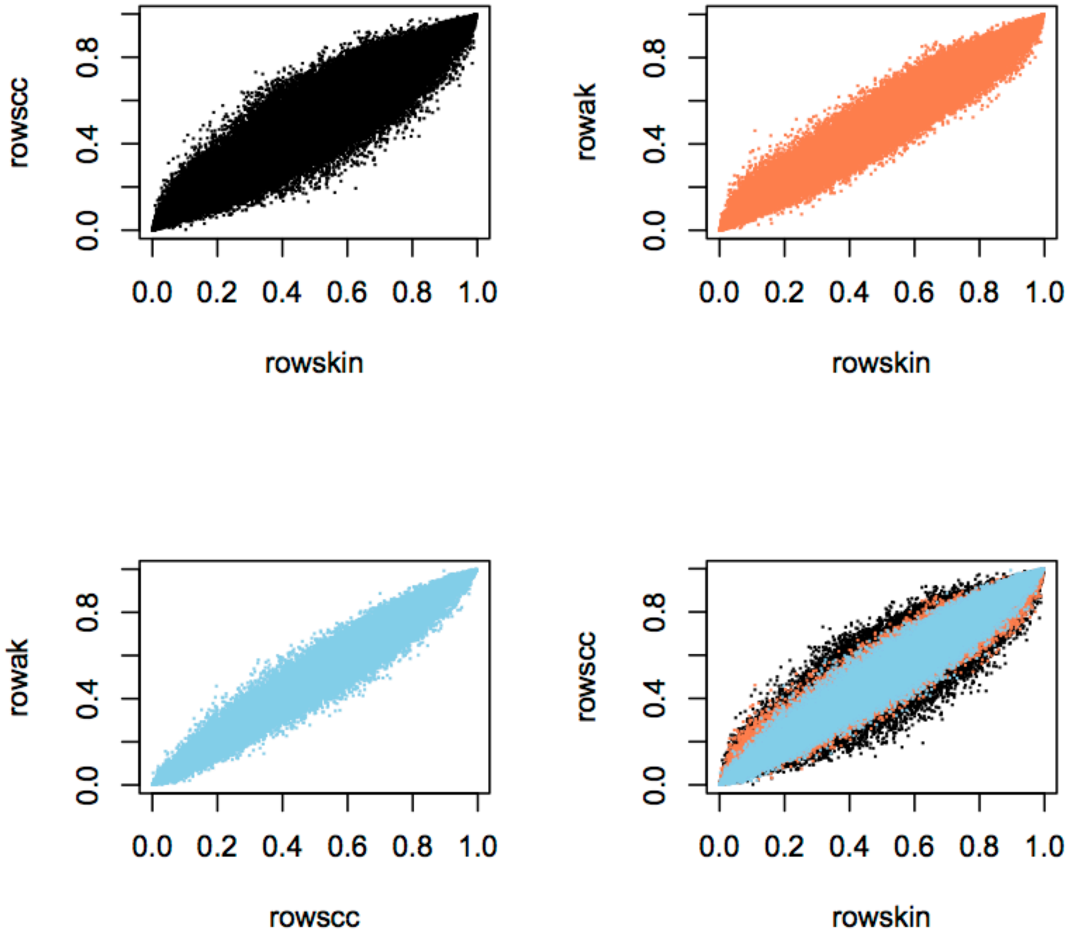


Figure 6.8 Correlation of mean methylation values in skin, AK and cSCC samples. Top left panel shows correlation of mean beta values of skin and cSCC in black, top right panel shows correlation of mean beta values in skin and AK in coral red, and bottom left panel shows correlation of mean beta values of cSCC and AK. Bottomright panel is a merge of previous three panels, showing the highest correlation between cSCC and AK, and the lowest between cSCC and skin.

Given that the number of DMPs detected in cSCC compared to AK was relatively low following p value adjustment, all comparisons are presented individually and individual cut-offs were used in each analysis.

6.3.4. Comparison of genome-wide methylation of normal skin and actinic keratosis reveals 1822 hypermethylated and 2372 hypomethylated genes

The 20 skin samples and 19 AK samples originated in a total of 20 patients, and a total of 14 patients contributed paired skin and an AK sample. The cut-off for differentially methylated probes comparing skin with AK was set at <0.01 (adjusted p value). The most stringent p value (<0.0001) revealed only 2 differentially methylated probes: for HOXB13 and PARD3B genes.

P value <0.01 identifies 8697 unique differentially methylated probes in AK. Of those probes, 4124 (47%) are hypomethylated in AK, and 4564 (53%) are hypermethylated in AK compared to normal skin (Figure 6.9). Heatmap of beta values corresponding to this probeset in all AK and skin samples is shown in Figure 6.10.

Hypomethylated probes translated into 1822 unique genes, while hypermethylated probes represented 2372 unique genes. 291 genes appeared in both lists due to the presence of multiple probes corresponding to various regions of the gene on the array. Gene set enrichment analysis of hyper- and hypomethylated genes was then conducted separately.

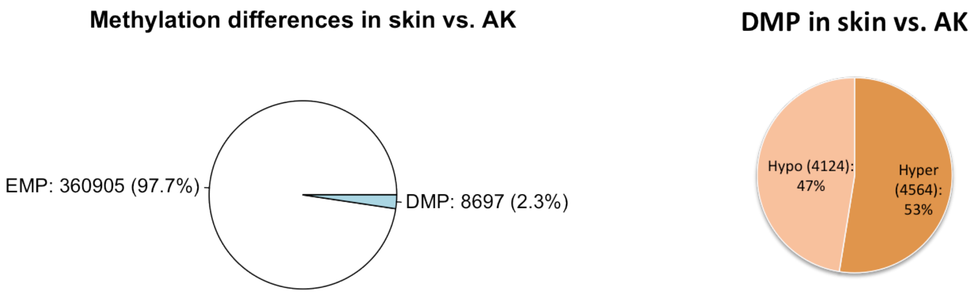


Figure 6.9 Percentage of differentially methylated probes (DMP) in AK compared to skin. EMP=equally methylated probes. Of differentially methylated probes (DMPs), 48% are hypermethylated in AK, 52% are hypomethylated compared to normal skin.

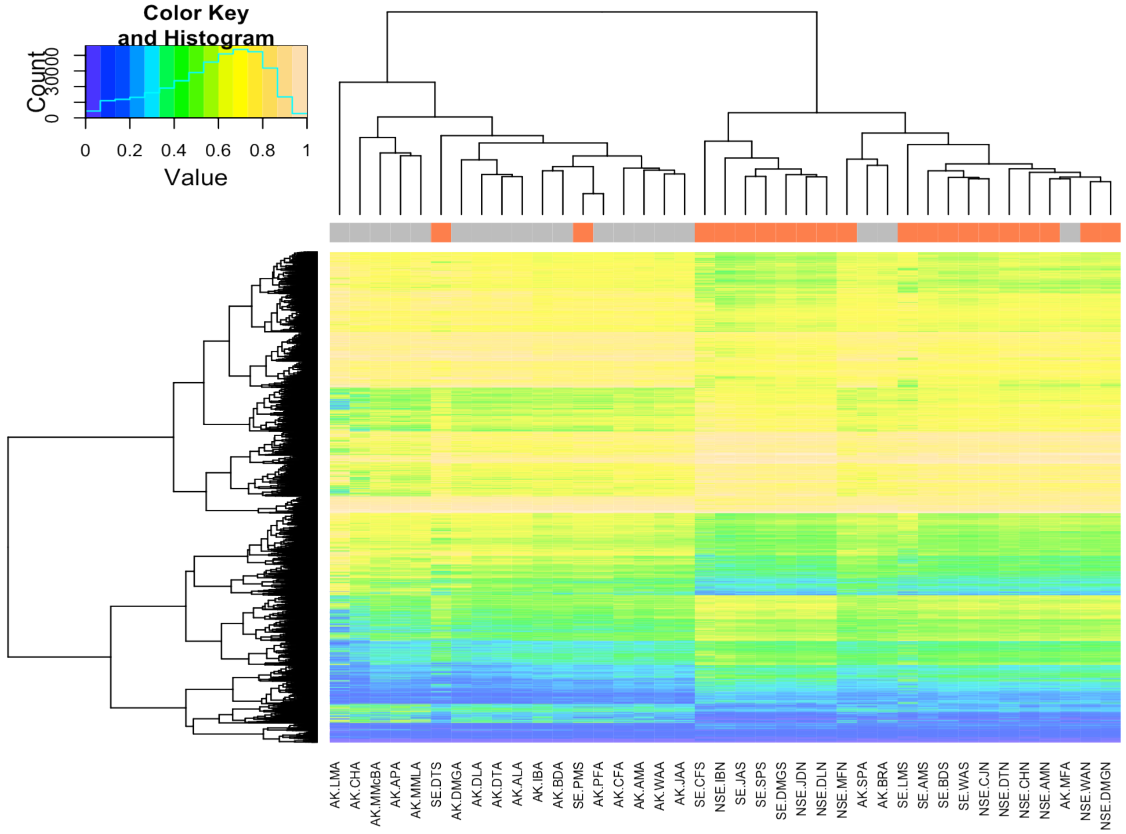


Figure 6.10 Heatmap of all differentially methylated probes (<0.01) in AK (grey) compared to normal skin (coral red). With the exception of 2 skin samples and 3 AKs, all samples clustered based on their tissue characteristics using this probseset.

Genomic distribution of DMP shows their presence on all somatic chromosomes, and indicates that the differences in mean beta values are evenly distributed across the chromosomes (Figure 6.11).

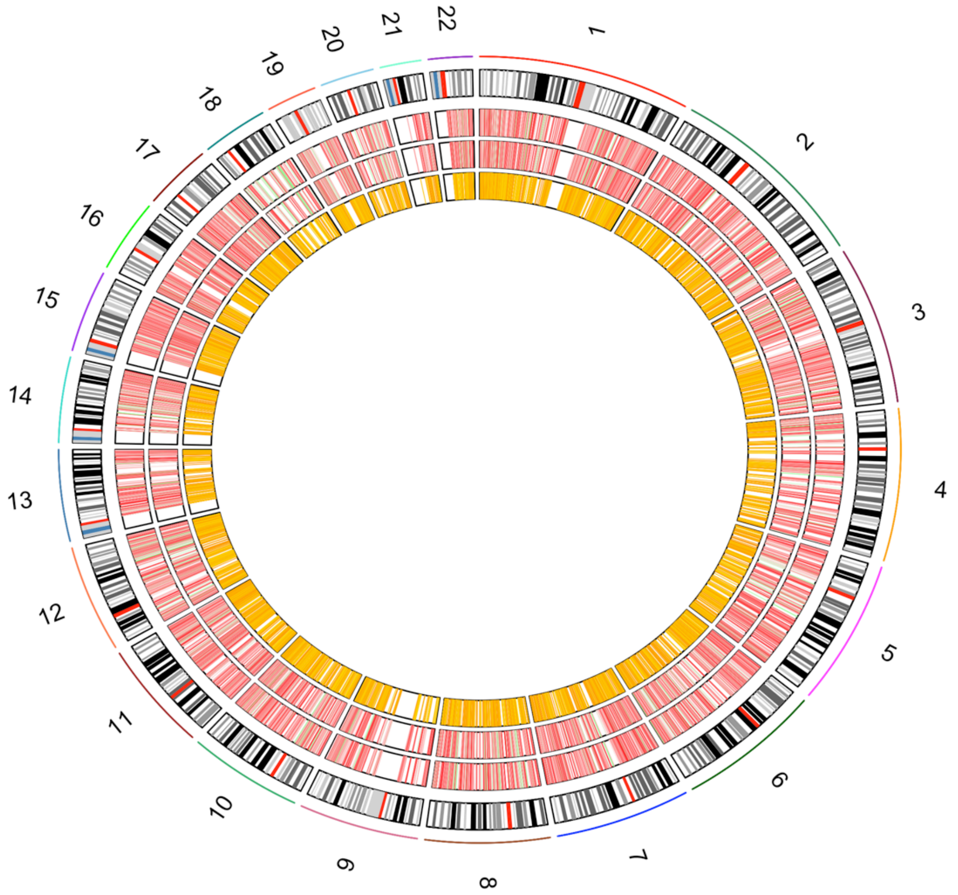


Figure 6.11 Genomic distribution of probes differentially methylated in AK. Outer circle=ideogram of individual chromosomes. Second outer cicle=mean beta values in skin, second inner circle=mean beta values in AK, inner circle=differences between mean values. For the red and green circles, green=0, white = 0.5 red=1 (beta value). For the inner circle, -0.35=yellow, 0.25=red.

6.3.5. Functional genomic distribution of differentially methylated probes in normal skin versus AK

Illumina 450K array classifies probes the following way: promoter probes, probes corresponding to intergenic sequences (Intergenic), gene body (Body) or 3'-UTR. Many probes have multiple functional annotations.

Functional genomic distribution of hypermethylated and hypomethylated probes presented in Figure 6.12 and Table 6.3 revealed differences in the relative representation of probes corresponding to intergenic regions, which were far more commonly hypomethylated, in addition to gene body methylation, 3'UTR and promoters that were more commonly hypermethylated in AK.

skin vs ak	Hyper	Hypo	P value
Intergenic	952	1552	<0.0001
3'UTR	554	196	<0.0001
Body	3876	2468	<0.0001
Promoter	2551	1794	<0.0001

Table 6.3 Functional annotation of hyper- and hypo-methylated probes. P-value based on chi-square test.

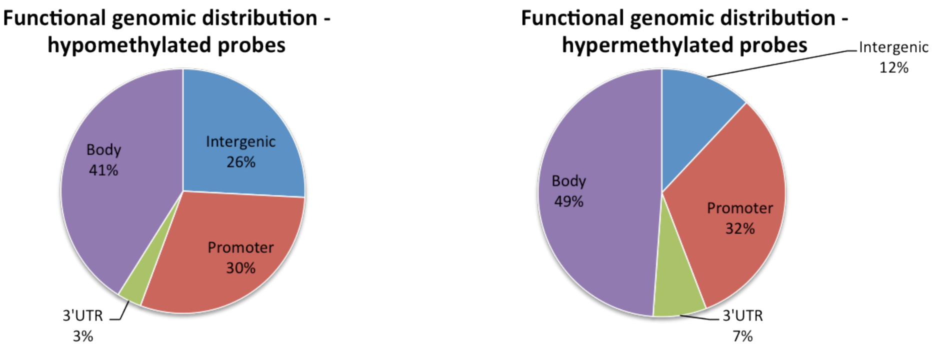


Figure 6.12 Functional genomic distribution of hyper- and hypo-methylated genes. Hypomethylation of probes corresponding to intergenic regions is far more common compared with hypermethylation.

6.3.6. CpG islands and the neighbourhood context

I then looked at the relative representation of probes classified based on their positional relationship to CpG islands. Hypomethylated probes corresponded more commonly to "open sea" regions (Figure 6.13, p value <0.0001).

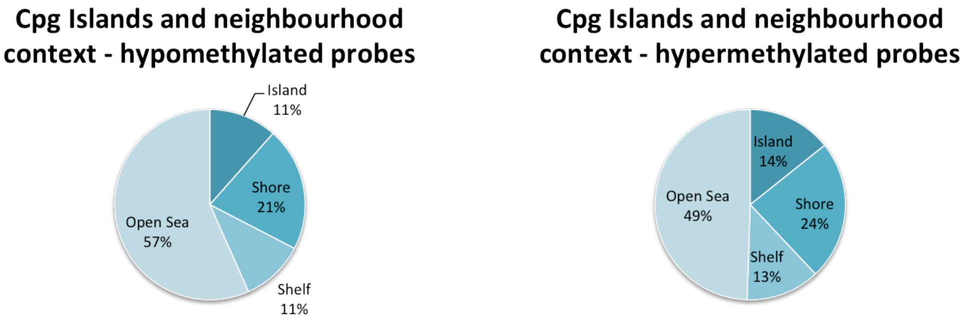


Figure 6.13 Probes corresponding to CpG islands and neighbourhood regions.

6.3.7. Pathway analysis of genes differentially hypomethylated in normal skin versus AK reveals dysregulation of cancer-related and metabolic pathways

The results of KEGG pathway analysis of hypomethylated genes revealed 84 dysregulated pathways (full list is presented in Appendix 10). Pathways with more than 20 annotated genes are shown in Figure 6.14. Selected pathways directly revelant to the oncogenic process are shown in Table 6.4.

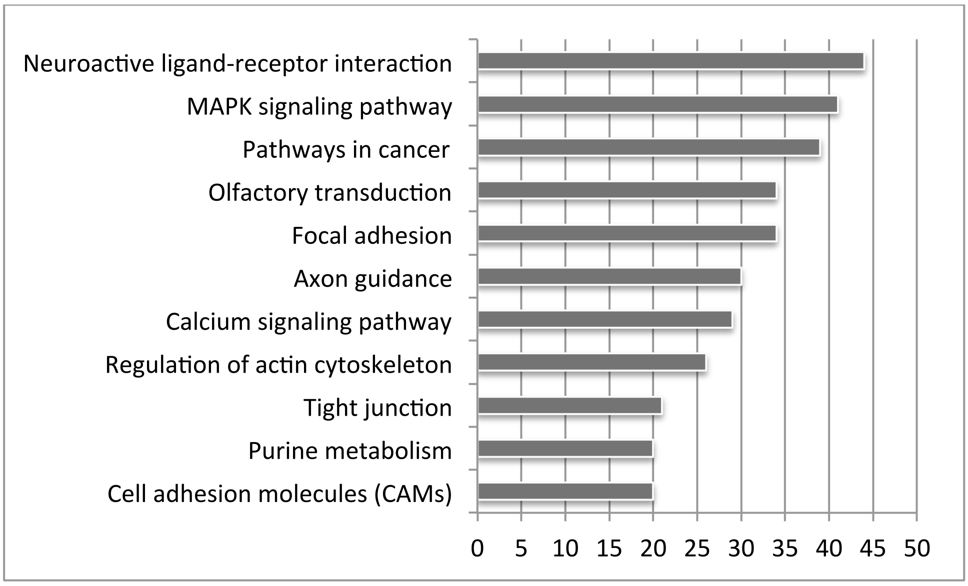


Figure 6.14 KEGG Pathways with most annotated hypomethylated genes (skin versus AK).

KEGG Pathway	Annotated genes	p-value	Adjusted p value
MAPK signaling pathway	41	3.84E-14	2.38E-12
Pathways in cancer	39	4.56E-10	1.06E-08
Tight junction	21	4.20E-08	7.10E-07
Focal adhesion	34	3.35E-13	1.56E-11
ErbB signaling pathway	16	1.76E-07	2.73E-06
Cell adhesion molecules (CAMs)	20	1.99E-07	2.85E-06
Regulation of actin cytoskeleton	26	2.70E-07	3.59E-06
Melanogenesis	16	1.64E-06	1.91E-05
Adherens junction	11	1.27E-04	7.88E-04
Gap junction	12	1.62E-04	9.74E-04
Wnt signaling pathway	16	2.31E-04	1.30E-03
mTOR signaling pathway	8	7.52E-04	3.50E-03
Endometrial cancer	7	3.49E-03	1.12E-02
Cell cycle	12	3.75E-03	1.18E-02
Non-small cell lung cancer	7	4.33E-03	1.34E-02
Pancreatic cancer	8	5.16E-03	1.55E-02

Renal cell carcinoma	8	5.16E-03	1.55E-02
Hedgehog signaling pathway	7	5.31E-03	1.57E-02
TGF-beta signaling pathway	9	5.58E-03	1.61E-02
Notch signaling pathway	6	8.68E-03	2.27E-02

Table 6.4 Selected dysregulated KEGG pathways. Dysregulation determined by overrepresentation in hypomethylated genes (skin versus AK).

6.3.8. Positional gene signature reveals overrepresentation of hypomethylated probes in three or more bands on chromosomes 5, 6 and 16

Positional gene signature of hypomethylated genes presented in Table 6.5 revealed 30 cytobands with overrepresentation of hypomethylated probes. Most common chromosomes were chromosome 6 (5 bands), 5 (4 bands) and 16 (3 bands).

Chromosomal band	Genes in band	Adjusted p value
chr10q25	9	1.01E-02
chr10q26	14	4.25E-02
chr11p15	41	6.89E-05
chr12q24	22	2.70E-02
chr15q21	12	2.70E-02
chr16p12	12	3.34E-02
chr16p13	29	2.04E-03
chr16q24	12	3.39E-02
chr1p36	51	1.22E-07
chr1q32	16	4.66E-02
chr20q11	14	3.51E-02
chr20q13	24	1.17E-03
chr21q22	23	5.28E-03
chr3p14	9	1.72E-02
chr3q21	18	1.77E-04
chr4p16	22	6.89E-05
chr4q34	5	2.91E-02
chr5p13	11	4.79E-03
chr5p14	6	2.03E-02
chr5p15	16	4.79E-03
chr5q35	15	4.28E-02
chr6p21	36	2.19E-02
chr6q16	6	3.39E-02
chr6q22	12	6.82E-03
chr6q25	10	3.14E-02
chr6q26	4	3.39E-02
chr7q21	14	4.70E-03
chr7q32	9	2.70E-02

chr8q24	24	1.77E-04
chr9q34	26	7.68E-04

Table 6.5 Positional signature of genes hypomethylated in AK.

Band 8q23 which has been shown to contain loss of genomic material in AK (see Chapter 3) did not overlap with significant loci containing hypomethylated genes.

6.3.9. Pathway analysis of hypermethylated genes reveals dysregulation of additional metabolic and oncogenic pathways

KEGG Pathway analysis of hypermethylated genes found significant dysregulation of 84 pathways (full list is presented in Appendix 11). Table 6.6 lists pathways directly relevant in cancer; Figure 6.15 shows pathways with 20 or more annotated genes.

Description	Annotated genes	p-value	Adjusted p value
Thyroid cancer	7	4.70E-04	1.51E-03
p53 signaling pathway	8	2.21E-02	4.83E-02
Melanoma	10	2.82E-03	8.07E-03
VEGF signaling pathway	10	4.65E-03	1.29E-02
Endometrial cancer	11	4.50E-05	1.86E-04
mTOR signaling pathway	11	4.50E-05	1.86E-04
Notch signaling pathway	12	2.54E-06	1.97E-05
Non-small cell lung cancer	13	2.01E-06	1.62E-05
Hedgehog signaling pathway	13	3.12E-06	2.23E-05
Colorectal cancer	13	1.03E-05	6.01E-05
ErbB signaling pathway	16	6.24E-06	4.00E-05

Table 6.6 Selected cancer-relevant KEGG pathways dysregulated in genes hypermethylated in AK.

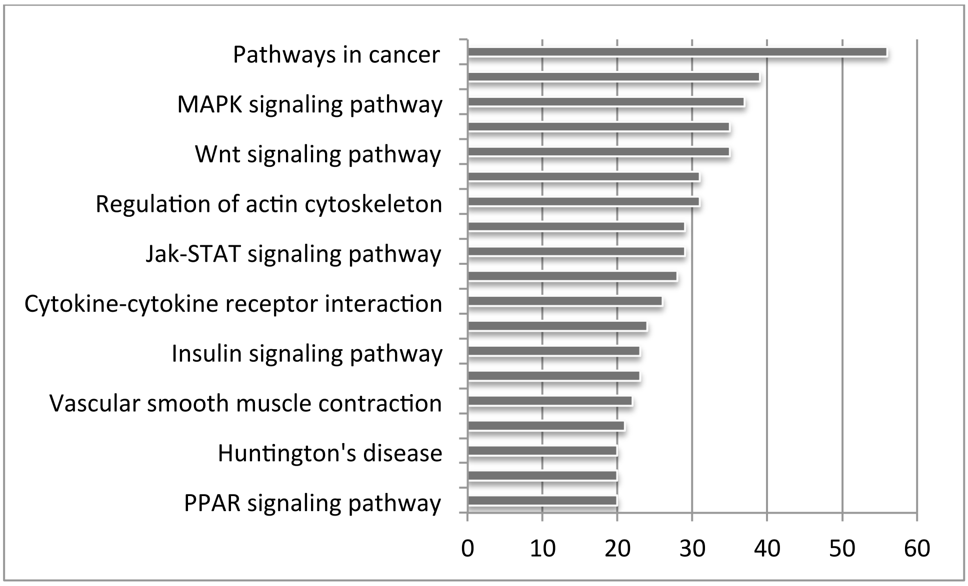


Figure 6.15 Dysregulated KEGG Pathways with 20 or more annotated genes hypermethylated in AK.

6.3.10. KEGG Pathways dysregulated in both hypo- and hypermethylated genes include critical oncogenic pathways

Overlapping KEGG pathways included many pathways directly relevant to skin cancer, including MAPK, Notch, TGF-beta, ErbB signalling pathway. Interestingly, Jak-STAT, VEGF, PPAR and p53 pathways were significantly dysregulated among hypermethylated genes only (highlighted in Table 6.7).

Shared KEGG Pathways	**Hypomethylated**	**Hypermethylated**
Acute myeloid leukemia	ABC transporters	Aminoacyl-tRNA biosynthesis
Adherens junction	Aldosterone-regulated sodium reabsorption	Arachidonic acid metabolism
Adipocytokine signaling pathway	Amino sugar and nucleotide sugar metabolism	B cell receptor signaling pathway
Amyotrophic lateral sclerosis (ALS)	Cardiac muscle contraction	Biosynthesis of unsaturated fatty acids
Apoptosis	Complement and coagulation cascades	Circadian rhythm - mammal
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Dorso-ventral axis formation	Colorectal cancer
Axon guidance	Drug metabolism - other enzymes	Ether lipid metabolism
Basal cell carcinoma	Fructose and mannose metabolism	Glycerolipid metabolism
Calcium signaling pathway	Galactose metabolism	Glycerophospholipid metabolism
Cell adhesion molecules (CAMs)	Glycolysis / Gluconeogenesis	Glycosaminoglycan biosynthesis - chondroitin sulfate
Cell cycle	Glyoxylate and dicarboxylate metabolism	Homologous recombination
Chemokine signaling pathway	Graft-versus-host disease	Huntington's disease
Chronic myeloid leukemia	Leishmania infection	Inositol phosphate metabolism
Cytokine-cytokine receptor interaction	Long-term depression	**Jak-STAT signaling pathway**
Dilated cardiomyopathy	Natural killer cell mediated cytotoxicity	Lysine degradation
ECM-receptor interaction	Olfactory transduction	Lysosome

Endocytosis	Pantothenate and CoA biosynthesis	Other glycan degradation
Endometrial cancer	Pathogenic Escherichia coli infection	**PPAR signaling pathway**
ErbB signaling pathway	Peroxisome	RNA degradation
Fc epsilon RI signaling pathway	Phenylalanine metabolism	Taurine and hypotaurine metabolism
Fc gamma R-mediated phagocytosis	Proximal tubule bicarbonate reclamation	Thyroid cancer
Focal adhesion	Pyruvate metabolism	Toll-like receptor signaling pathway
Gap junction	Riboflavin metabolism	Ubiquitin mediated proteolysis
Glioma	Ribosome	**VEGF signaling pathway**
GnRH signaling pathway	SNARE interactions in vesicular transport	Valine, leucine and isoleucine biosynthesis
Hedgehog signaling pathway	Systemic lupus erythematosus	Vasopressin-regulated water reabsorption
Hematopoietic cell lineage	Type I diabetes mellitus	Vibrio cholerae infection
Histidine metabolism	Type II diabetes mellitus	alpha-Linolenic acid metabolism
Hypertrophic cardiomyopathy (HCM)	Tyrosine metabolism	**p53 signaling pathway**
Insulin signaling pathway		
Leukocyte transendothelial migration		
Long-term potentiation		
MAPK signaling pathway		
Melanogenesis		
Melanoma		
Neuroactive ligand-receptor interaction		
Neurotrophin signaling pathway		
Non-small cell lung cancer		
Notch signaling pathway		
O-Glycan biosynthesis		
Oocyte meiosis		
Pancreatic cancer		
Pathways in cancer		
Phosphatidylinositol signaling system		
Progesterone-mediated oocyte maturation		
Prostate cancer		
Purine metabolism		
Regulation of actin cytoskeleton		
Renal cell carcinoma		
Small cell lung cancer		
T cell receptor signaling pathway		
TGF-beta signaling pathway		
Tight junction		
Vascular smooth muscle contraction		
Wnt signaling pathway		
mTOR signaling pathway		

Table 6.7 Dysregulated pathways in both hypo- and hyper-methylated genes in AK. Overlapping pathways are listed in the first column.

6.3.11. Positional gene signature of hypermethylated genes finds overrepresentation in three or more bands on chromosomes 1, 6, 10, 16, 17

Positional gene signature of hypermethylated genes found 31 chromosomal bands to contain overrepresented hypermethylated genes listed in Table 6.8. Of those 31 bands, 14 also contained hypomethylated genes.

Gene Set Name	# Genes in Overlap (k)	p-value	FDR q-value
chr10q22	17	8.07E-04	1.14E-02
chr10q25	10	7.58E-04	1.12E-02
chr10q26	19	4.92E-04	8.01E-03
chr11p15	40	5.05E-04	8.01E-03
chr11q13	47	7.68E-10	8.35E-08
chr12q13	29	1.11E-03	1.45E-02
chr12q24	30	1.19E-04	2.58E-03
chr13q34	8	4.72E-03	4.96E-02
chr15q22	14	2.44E-03	2.75E-02
chr16p13	43	2.59E-08	2.11E-06
chr16q22	24	3.97E-06	1.44E-04
chr16q24	15	1.32E-03	1.66E-02
chr17q11	18	5.16E-04	8.01E-03
chr17q21	37	9.20E-06	2.73E-04
chr17q25	31	1.86E-07	8.67E-06
chr19p13	51	1.05E-03	1.43E-02
chr19q13	73	2.27E-04	4.11E-03
chr1p34	30	4.96E-08	3.24E-06
chr1p36	72	1.68E-15	2.73E-13
chr1q32	26	8.49E-06	2.73E-04
chr2p23	15	1.76E-04	3.58E-03
chr3p14	13	1.78E-05	4.46E-04
chr3p21	35	3.42E-07	1.39E-05
chr4p16	18	2.36E-03	2.75E-02
chr5q35	19	1.76E-03	2.13E-02
chr6p21	76	9.48E-16	2.73E-13
chr6p25	9	4.09E-03	4.44E-02
chr6q24	16	1.53E-07	8.32E-06
chr7p22	17	1.66E-05	4.46E-04
chr8q24	25	8.26E-05	1.92E-03
chr9q34	28	2.09E-04	4.01E-03

Table 6.8 Cytobands containing significantly more hypermethylated genes in AK.

Comparison with regions of genomic gain in AK detected with SNP arrays (details in Chapter 3) revealed overlap with 8q24.11, 8q24.12, 8q24.21, 9q34.3, and 11q13.3. Hypermethylated genes corresponding to 11q13.3 included TPCN2 (two pore segment channel 2) and MRGPRF (MAS-related GPR, member F). Genes corresponding to 9q34.3 included PMPCA (peptidase (mitochondrial processing) alpha), SNAPC4 (small nuclear RNA activating complex, polypeptide 4, 190kDa), SEC16A (SEC16 homolog A (S. cerevisiae)), EXD3 (exonuclease 3'-5' domain containing 3), EHMT1 (euchromatic histone-lysine N-methyltransferase 1), and CACNA1B (calcium channel, voltage-dependent, N type, alpha 1B subunit). Segment 8q24.21 codes for PVT1 (Pvt1 oncogene (non-protein coding)), and ASAP1 (ArfGAP with SH3 domain, ankyrin repeat and PH domain 1). 8q24.12 codes for EXT1 (exostosin 1) and TAF2 (TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa). 8q24.11 did not contain annotated hypermethylated genes.

Cytobands 1p36, 3p14 and 8q24 contained significantly overerepresented individual genes that were both hyper and hypomethylated. Bands that overlapped between both hyper and hypomethylated genes are shown in Table 6.9.

Shared bands	Hypo only	Hyper only
chr10q25	chr15q21	chr10q22
chr10q26	chr16p12	chr11q13
chr11p15	chr20q11	chr12q13
chr12q24	chr20q13	chr13q34
chr16p13	chr21q22	chr15q22
chr16q24	chr3q21	chr16q22
chr1p36	chr4q34	chr17q11
chr1q32	chr5p13	chr17q21
chr3p14	chr5p14	chr17q25
chr4p16	chr5p15	chr19p13
chr5q35	chr6q16	chr19q13
chr6p21	chr6q22	chr1p34
chr8q24	chr6q25	chr2p23
chr9q34	chr6q26	chr3p21
	chr7q21	chr6p25
	chr7q32	chr6q24
		chr7p22

Table 6.9 Shared bands with overrepresentation of hyper- and hypo-methylated genes. Bands that contain only hyper- or hypo-methylated genes are presented in additional columns.

6.3.12. Integration with expression data reveals generally poor correlation between gene expression and methylation

Mean log2 ratio of expression values of all hyper- and hypomethylated genes revealed generally lower expression levels in hypermethylated genes (mean log2 ratio of 0.013 in hypermethylated genes versus 0.097 in hypomethylated genes, Figure 6.16).

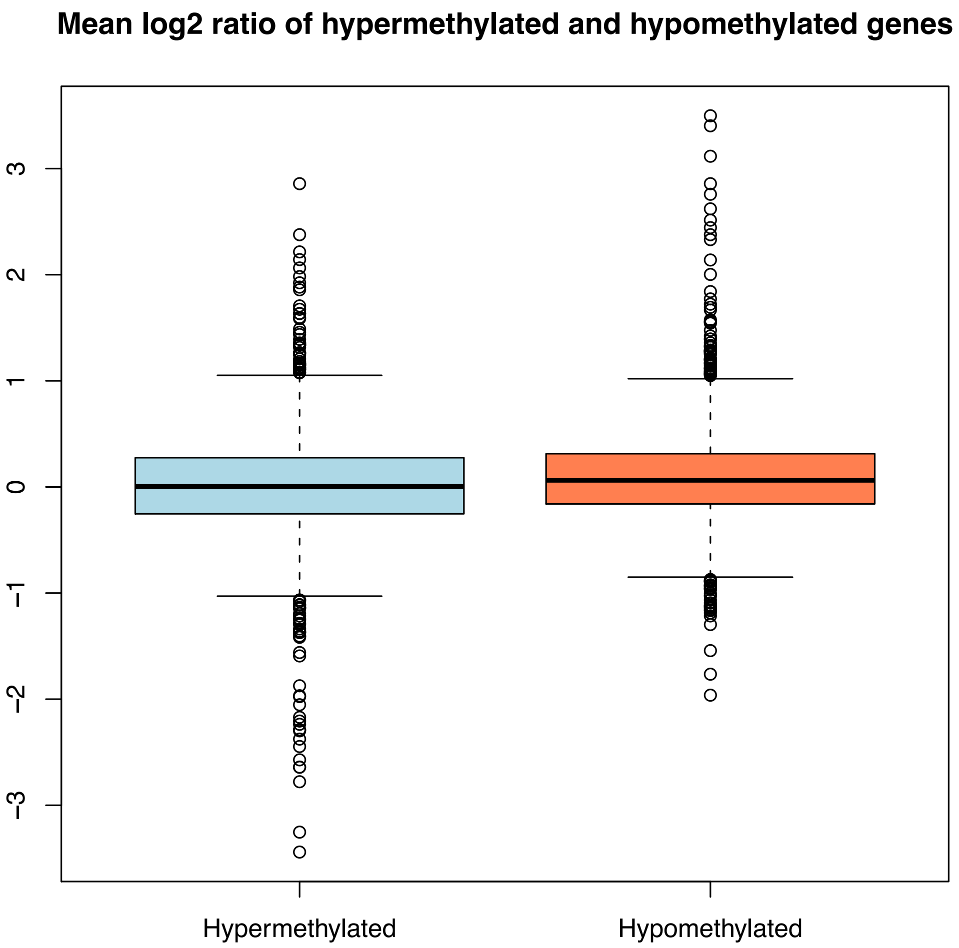
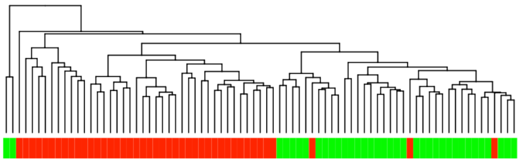
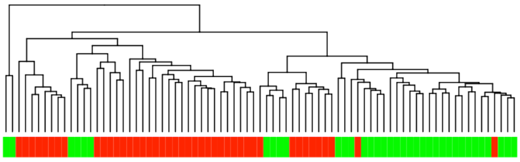


Figure 6.16 Mean log2 expression values of hyper- and hypomethylated genes in AK. The median of hypomethylated genes is higher compared to hypermethylated genes (p value<0.001).

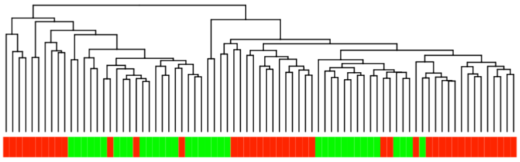
The expression levels of both hyper- and hypomethylated genes separates normal skin from AK into two major clusters, in contrast to the clustering obtained with a set of 1500 randomly selected genes (Figure 6.17).



Hypermethylated genes



Hypomethylated genes



Random 1500 genes

Figure 6.17 Clustering of normal skin (green) and AK samples (red) based on the expression levels of hyper-, hypomethylated genes and a set of random 1500 genes. Random rows were selected in R using the "sample" function.

I then explored differences in expression based on specific methylation values: all genes were grouped according to their methylation level into groups of more than 20% methylation difference, and then by 5% differences, creating the following groups: 15-20%, 10-15%, 5-10%, 0-5%, with negative sign for hypomethylated genes. As shown in Figure 6.18 and in the notched boxplot in Figure 6.20, expression levels generally increase with decreasing degree of methylation, yet this trend is not linear as previously observed (see Chapter 5).

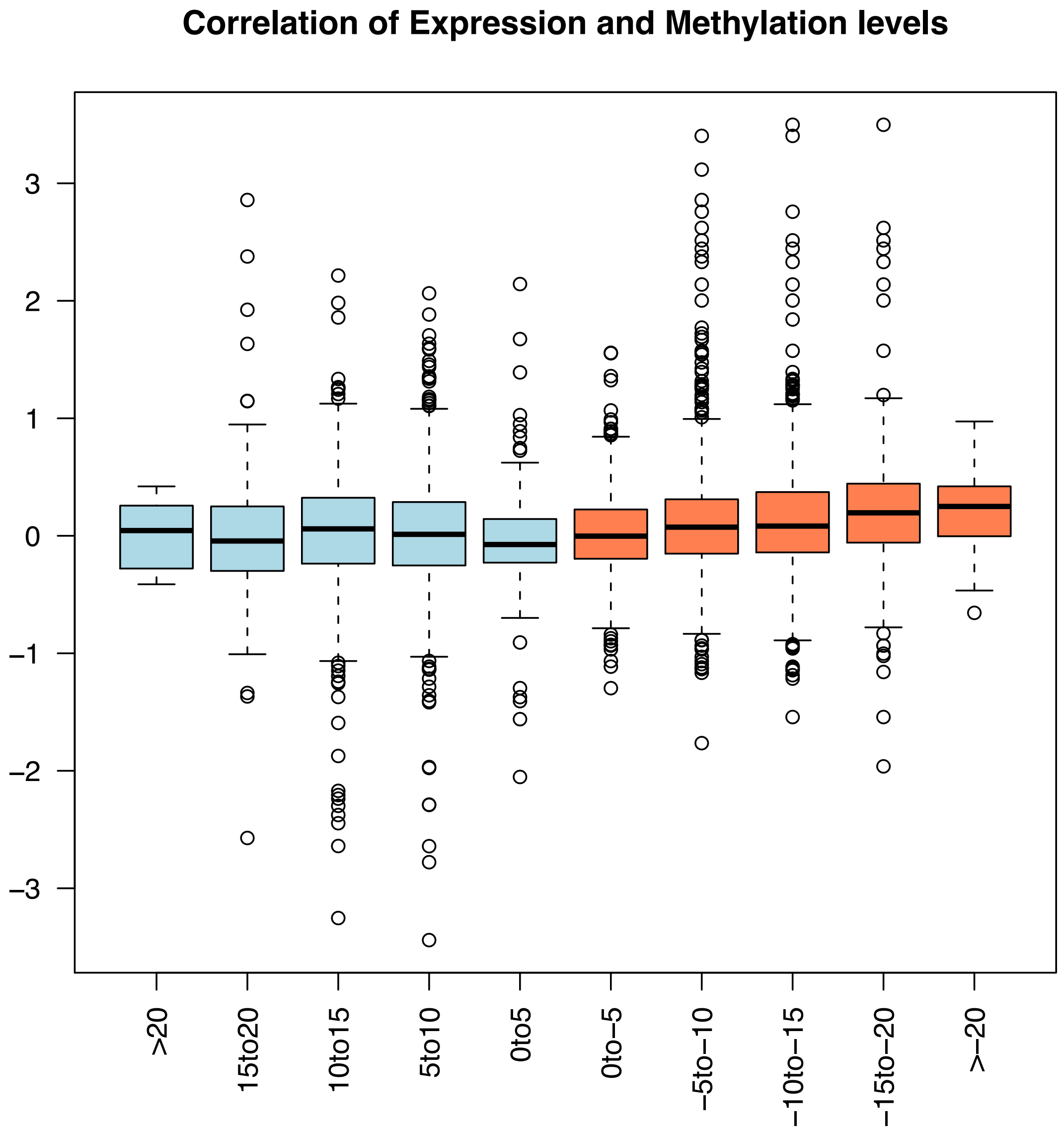


Figure 6.18 Mean log2 ratio of expression levels in AK correlated with different levels of methylation. Light blue=hypermethylated genes, coral=hypomethylated genes.

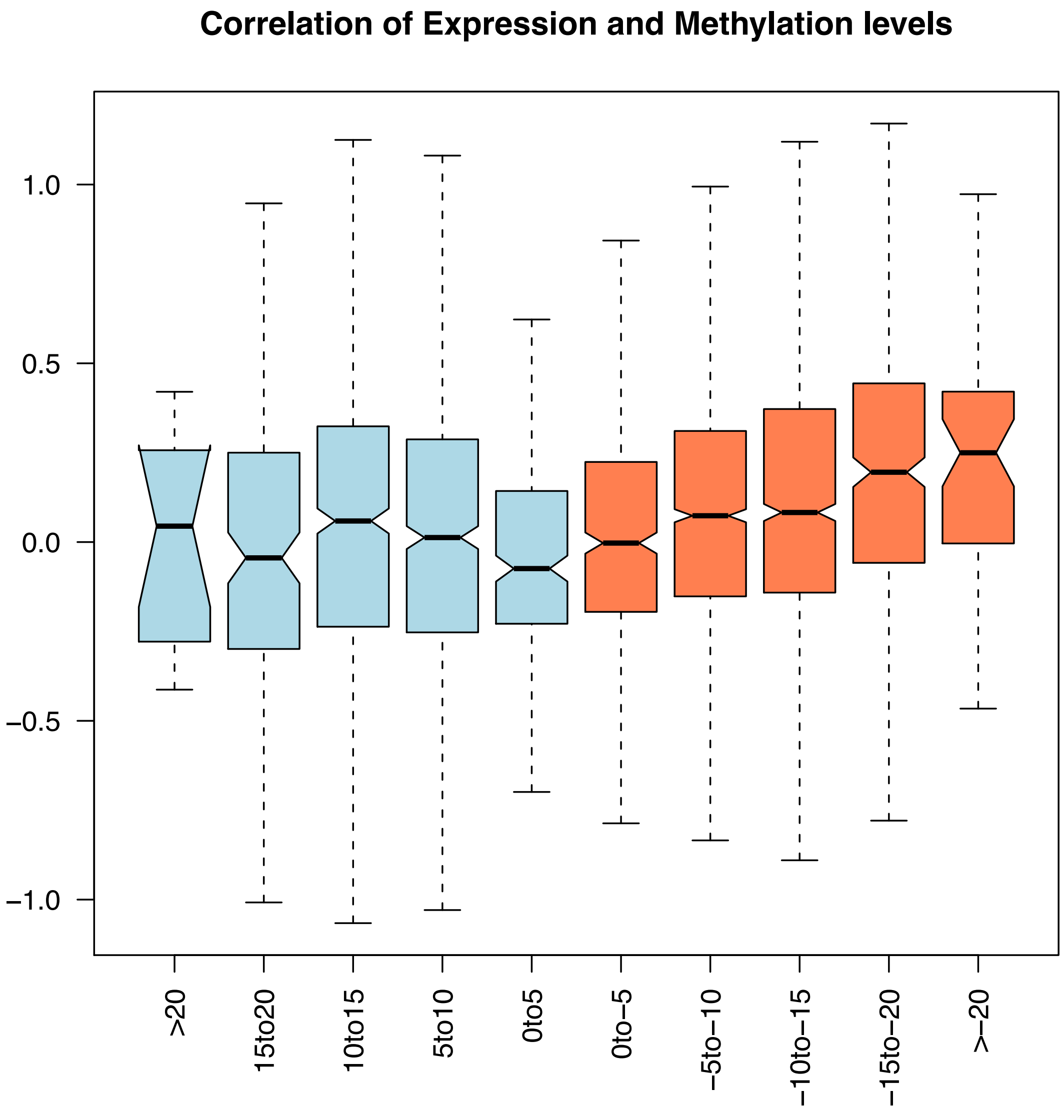


Figure 6.19 Mean log2 ratio of expression levels in AK correlated with different levels of methylation not depicting the outliers. Light blue=hypermethylated genes, coral=hypomethylated genes.

I then hypothesised that character of probes with respect to functional genomic distribution (promoter, gene body, 3'UTR) may impact expression.

Indeed, as shown in Figure 6.20 and Figure 6.21, hypermethylation of 3'U regions leads to increased gene expression, while promoter and gene body methylation decreases transcription. All within-group comparisons (3'UTR hypermethylated vs. 3'UTR hypomethylated etc.) were statistically significant (p value <0.05).

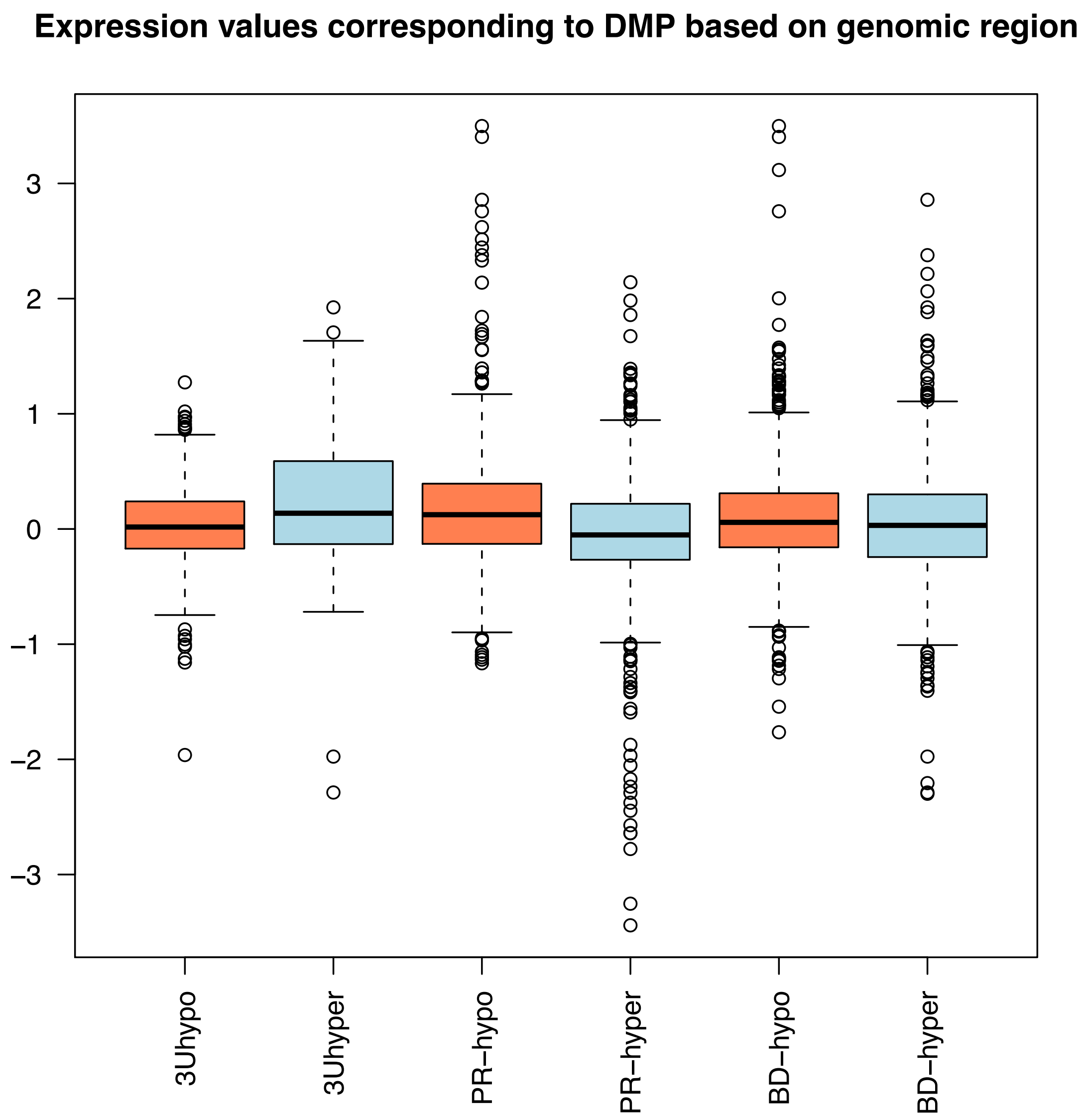


Figure 6.20 Mean log2 expression values based on functional genomic distribution of probes. Hypermethylated probes corresponding to 3'UTR lead to increased expression, while hypermethylated probes corresponding to gene bodies (BD) and promoters (PR) decrease expression.

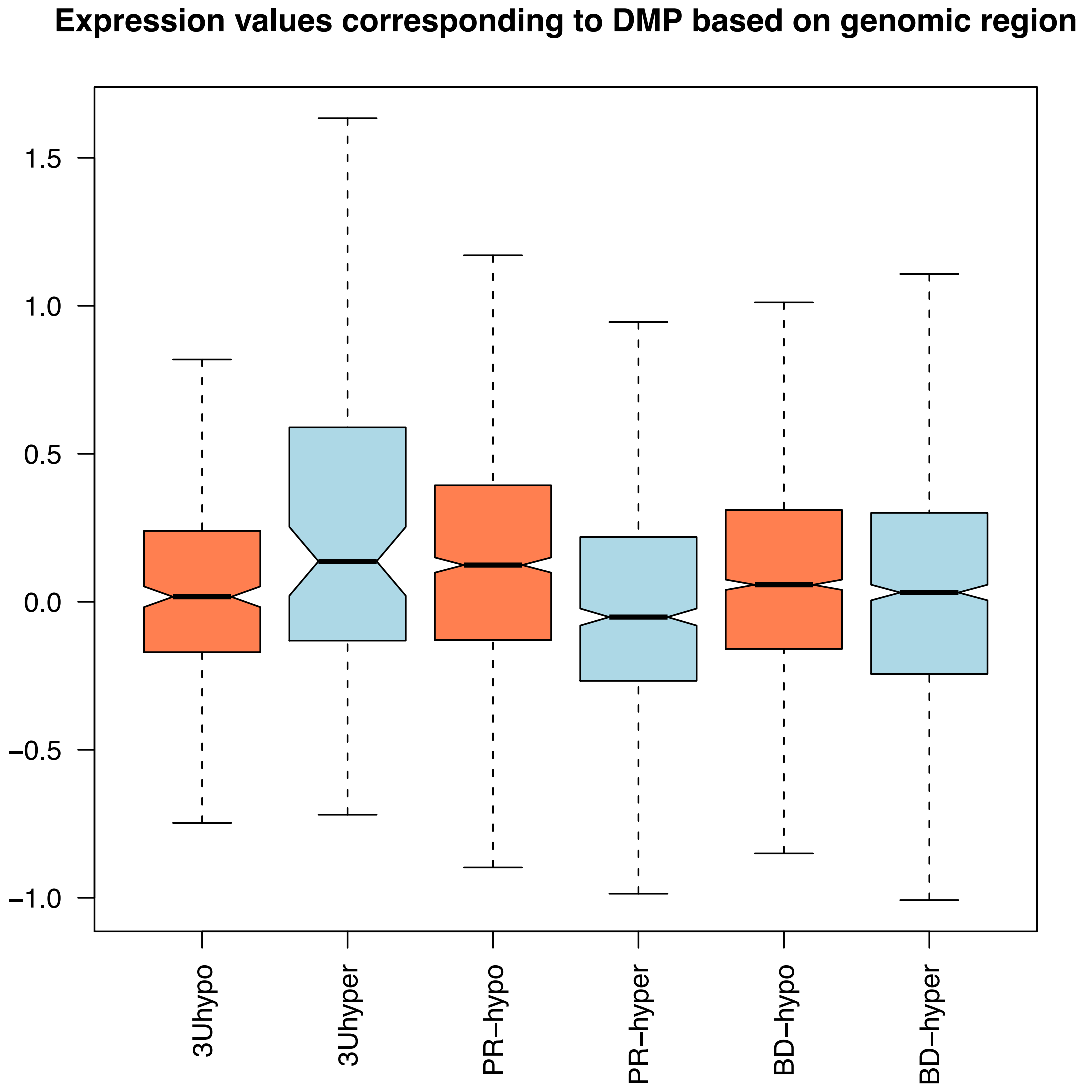


Figure 6.21 Mean log2 expression values based on functional genomic distribution of probes (outliers not depicted). Hypermethylated probes corresponding to 3'UTR lead to increased expression, while hypermethylated probes corresponding to gene bodies (BD) and promoters (PR) decrease expression.

6.3.13. Enrichment for transcription factor binding sites finds possible association between FOXF2, PITX2, RUNX1 and SMAD3 and expression levels of their target genes

I then proceeded to determine potential transcription factors involved in AK onset by conducting enrichment analysis of TF-binding motifs in hyper- and hypomethylated promoters, and then a regression analysis of expression levels of target genes. Motif analysis calculating enrichment of transcription factor binding sites in hypermethylated and hypomethylated promoters using the top 100 enriched sequences revealed an overlap of motifs in 52 instances. Overlapping motifs (enriched for in both hyper- and hypomethylated promoters) are listed in Appendix 12 along with targeting transcription factors (TF). Motifs unique for genes with hypomethylated and hypermethylated promoters and respective TFs are listed in Appendices 13 and 14, respectively.

Unique TF targeting both hyper- and hypomethylated promoters are listed in Table 6.10. This list contains several known oncogenes (JUN, ETS2), and TF strongly associated with cancer (LEF1, PITX2, GATA1), or stemness (TCF3, NFATC1), in addition to homeobox genes (POU2F1, POU3F2, POU6F1, LHX3), and forkhead DNA-binding proteins (FOXA1, FOXC1, FOXF2). FOXF2, JUN, PITX2, TCF3 (upregulated), SP1, GATA3 (downregulated) also demonstrated significant dysregulation at a transcriptional level (p value<0.01).

TF	Name
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
DBP	D site of albumin promoter (albumin D-box) binding protein
ESRRA	estrogen-related receptor alpha
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
FOXA1	forkhead box A1
FOXC1	forkhead box C1
FOXF2	forkhead box F2
GATA1	GATA binding protein 1 (globin transcription factor 1)
GATA3	GATA binding protein 3
GFI1	growth factor independent 1
IRF1	interferon regulatory factor 1
JUN	jun oncogene
LEF1	lymphoid enhancer-binding factor 1
LHX3	LIM homeobox 3
MAZ	MYC-associated zinc finger protein (purine-binding transcription factor)
MEF2A	MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
MEIS1	Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)
MLLT7	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 7
MYOD1	myogenic differentiation 1
NFAT	NFATC
PAX4	paired box gene 4
PITX2	paired-like homeodomain transcription factor 2
POU2F1	POU domain, class 2, transcription factor 1
POU3F2	POU domain, class 3, transcription factor 2
POU6F1	POU domain, class 6, transcription factor 1
REPIN1	replication initiator 1
SP1	Sp1 transcription factor
TAF	TATA Box Binding Protein
TCF3	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
TCF8	transcription factor 8 (represses interleukin 2 expression)
VSX1	visual system homeobox 1 homolog, CHX10-like (zebrafish)

Table 6.10 Transcription factors that target both hyper- and hypomethylated genes.

TF targeting genes with hypomethylated promoters in AK are listed in Table 6.11. This list contains additional homeobox genes (CRX, HOXA4, PRRX2), oncogenes (RUNX1) and forkhead DNA-binding proteins. CEBPG, FOXD1, FOXM1, NFIL3 and RUNX1 were significantly upregulated in AK based on expression array data (p value<0.01), while NR3C1, RXRB, SMAD3, STAT5A, STAT5B and TEF were significantly down-regulated in AK.

TF	Name
CBFA2T2	core-binding factor, runt domain, alpha subunit 2; translocated to, 2 CBFA2T3
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
CEBPG	CCAAT/enhancer binding protein (C/EBP), gamma
CRX	cone-rod homeobox
CUTL1	cut-like 1, CCAAT displacement protein (Drosophila)
FOXD1	forkhead box D1
FOXJ1	forkhead box J1
FOXJ2	forkhead box J2
FOXM1	forkhead box M1
FOXO1A	forkhead box O1A (rhabdomyosarcoma)
FOXQ1	forkhead box Q1
GABPA	GA binding protein transcription factor, alpha subunit 60kDa GABPB2
GCM1	glial cells missing homolog 1 (Drosophila)
HOXA4	homeobox A4
NFE2L1	nuclear factor (erythroid-derived 2)-like 1 MAFG
NFE2L2	nuclear factor (erythroid-derived 2)-like 2
NFIL3	nuclear factor, interleukin 3 regulated
NKX2-5	NK2 transcription factor related, locus 5 (Drosophila)
NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
PRRX2	paired related homeobox 2
RUNX1	runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)
RXRB	retinoid X receptor, beta
SMAD3	SMAD, mothers against DPP homolog 3 (Drosophila)
SRF	serum response factor (c-fos serum response element-binding transcription factor)
STAT5A	signal transducer and activator of transcription 5A
STAT5B	signal transducer and activator of transcription 5B
TCF11 MAFG	v-maf musculoaponeurotic fibrosarcoma oncogene homolog G (avian)
TCF1	transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
TEF	thyrotrophic embryonic factor

Table 6.11 Transcription factors that target promoters of hypomethylated genes.

Enriched TF that target genes with hypermethylated promoters are listed in Table 6.12. KLF12 was significantly upregulated, AR, TEAD1 and ZNF238 were significantly down-regulated in AK. The remainder of TF was either not present on the expression array or their expressen level did not change in the progression from normal skin to AK.

TF	Name
AR	androgen receptor (dihydrotestosterone receptor; testicular feminization; spinal and bulbar muscular atrophy; Kennedy disease)
ATF3	activating transcription factor 3
ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)
CDC5L	CDC5 cell division cycle 5-like (S
ESR1	estrogen receptor 1
FOXO1A	forkhead box O1A (rhabdomyosarcoma)
GATA6	GATA binding protein 6
GTF3A	general transcription factor IIIA
HAND1	heart and neural crest derivatives expressed 1
HSF1	heat shock transcription factor 1
KLF12	Kruppel-like factor 12
LMO2	LIM domain only 2 (rhombotin-like 1)
NF1	neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)
NKX2-2	NK2 transcription factor related, locus 2 (Drosophila)
NR1H4	nuclear receptor subfamily 1, group H, member 4
PBX1	pre-B-cell leukemia transcription factor 1
PCAF	p300/CBP-associated factor
SPI1	spleen focus forming virus (SFFV) proviral integration oncogene spi1
SREBF1	sterol regulatory element binding transcription factor 1
SRF	serum response factor (c-fos serum response element-binding transcription factor)
TAL1	T-cell acute lymphocytic leukemia 1 TCF3
TCF12	transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)
TCF1	transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
TEAD1	TEA domain family member 1 (SV40 transcriptional enhancer factor)
TFAP4	transcription factor AP-4 (activating enhancer binding protein 4)
TITF1	thyroid transcription factor 1
TLX2	T-cell leukemia homeobox 2
ZIC3	Zic family member 3 heterotaxy 1 (odd-paired homolog, Drosophila)
ZNF238	zinc finger protein 238

Table 6.12 Transcription factors that target hypermethylated promoters.

I then hypothesised that regulation of differentially expressed transcription factors may be due to methylation changes in the genomic sequence of those factors. As shown in Table 6.13, 10/23 (43%) TF were significantly differentially methylated, and methylation status correlated with expected expression levels in all instances except for 3'UTR methylation of STAT5B, which showed direction of fold change opposite to expected based on the methylation status.

TF – hypermethylated promoters enrichment	Change in AK	methylation in AK	Site

CEBPG	up	x	
FOXD1	up	x	
FOXM1	up	x	
NFIL3	up	x	
RUNX1	up	hypomethylated	Promoter
NR3C1	down	x	
RXRB	down	hypermethylated	Body
SMAD3	down	hypermethylated	Promoter,Body
STAT5A	down	hypermethylated	Promoter
STAT5B	down	hypermethylated	3'UTR
TEF	down	X	
TF-hypomethylated promoters enrichment			
AR	down	X	
KLF12	up	hypomethylated	Body
TEAD1	down	hypermethylated	Body
ZNF238	down	hypermethylated	Promoter
TF-both			
FOXF2	up	X	
GATA3	down	hypermethylated	Body
JUN	up	X	
PITX2	up	hypomethylated	Promoter, Body
SP1	Down	X	
TCF3	up	X	

Table 6.13 Methylation status of transcription factors that target promoters of differentially methylated genes. X=methylation status not significantly different in AK.

I then explored the fold change of target genes of overexpressed TFs based on the hypothesis that functionally, target genes of overexpressed TFs with hypomethylated promoters would be up-regulated. Among target genes with hypomethylated promoters, 26/78 (33%) were upregulated (Figure 6.22). This is a higher proportion compared to genes with hypomethylated promoters that are targets of down-regulated TF (21/81, 26%) but this difference is not statistically significant (p value=0.4).

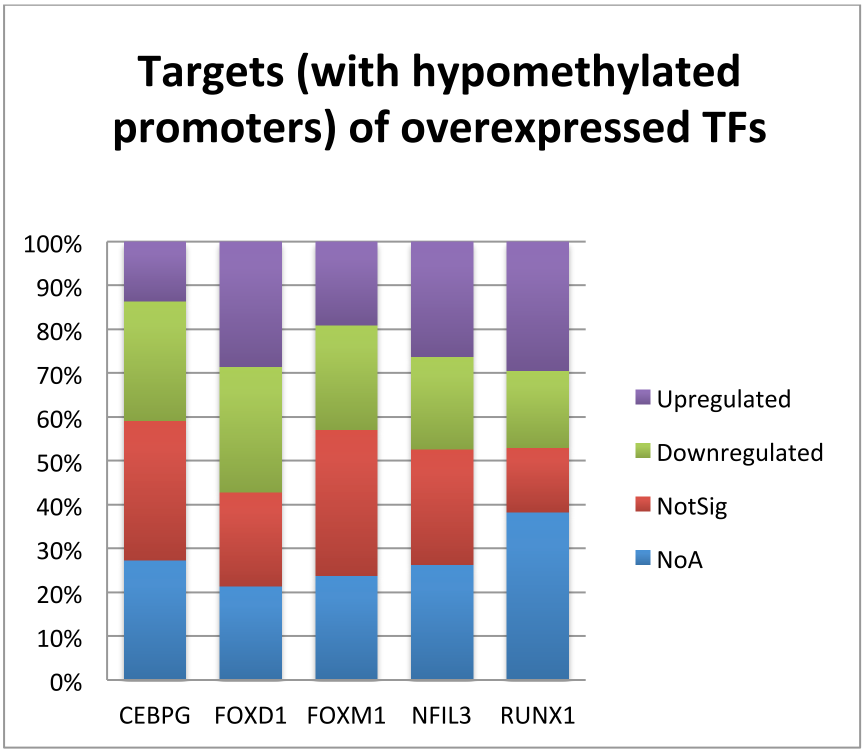


Figure 6.22 Proportion of upregulated and downregulated targets with hypomethylated promoters of overexpressed TFs. NotSig=expression levels available yet not significantly different in AK, NoA=gene not present on the expression array (expression levels not available). Targets of RUNX1 show the highest proportion of up-regulated genes.

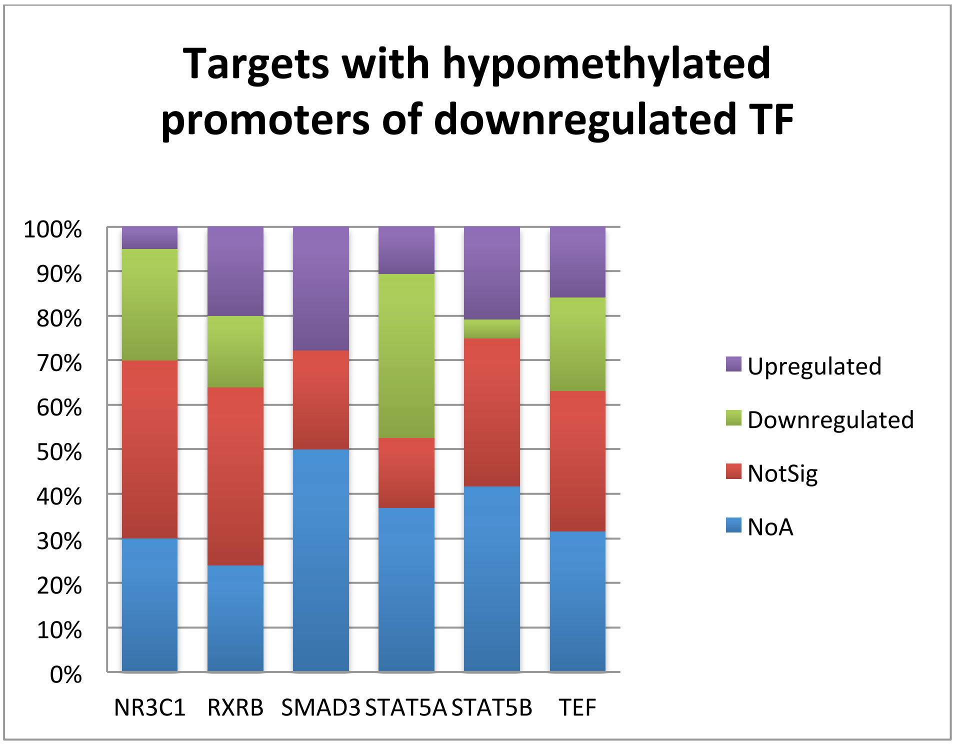


Figure 6.23 Cumulative percentage plot showing proportion of upregulated, downregulated targets with hypomethylated promoters of down-regulated TFs. NotSig=expression levels in AK not significantly different from normal skin, NoA=gene not present on the expression array and expression levels not available. The targets of SMAD3 that were significantly dysregulated were all overexpressed.

Among target genes with hypermethylated promoters, one target gene containing a binding motif for AR (down-regulated TF) was up-regulated (1/12, 8%) which was not significantly different from targets of down-regulated TF with hypomethylated promoters (p=0.14). Targets with hypermethylated promoters of up-regulated TF were upregulated in 11/52 instances (21%) but this was borderline not significant in comparison with targets with hypomethylated promoters targeted by up-regulated TF (p value =0.05).

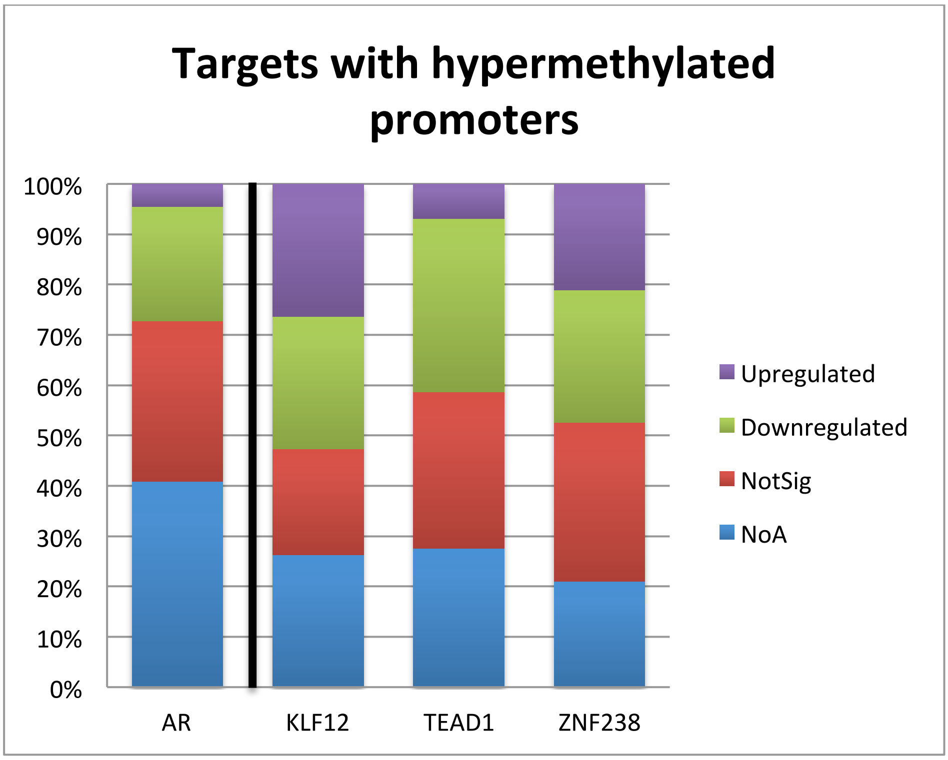


Figure 6.24 Targets with hypermethylated promoters. AR is down-regulated, the remaining TF are up-regulated. NotSig=expression levels in AK not significantly different from normal skin, NoA=gene not present on the expression array and expression levels not available.

Cumulative percentage graphs of target genes with hyper- and hypomethylated promoters are shown in Figure 6.25. Overexpression of target genes with hypomethylated promoters was more common (82/292, 28% versus 58/363, 16%), and this difference was statistically significant (p value=0.0003).

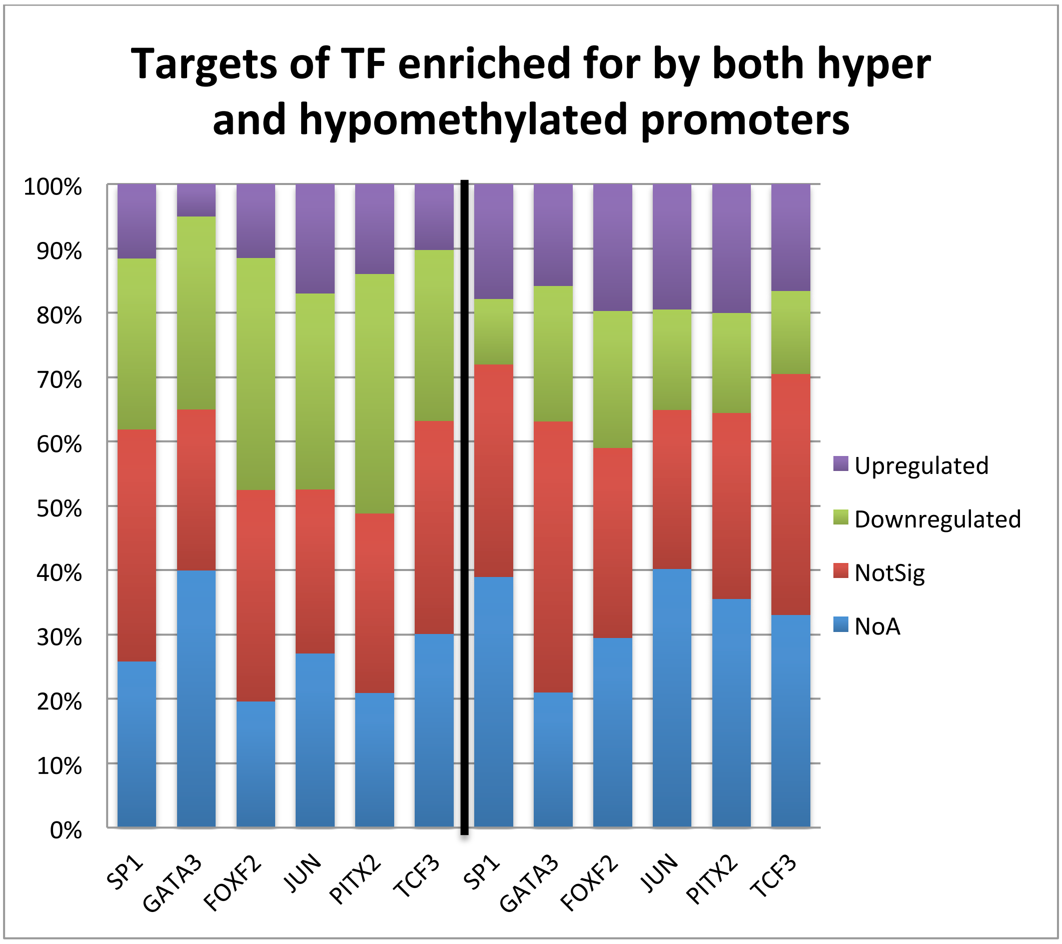


Figure 6.25 Cumulative percentage plot of genes that are targeted by TF that are enriched for by both hyper- and hypomethylated genes. Black line separates genes with hypermethylated promoters (left) from those with hypomethylated promoters (right).

Linear regression analysis assessing the association between fold-change of significantly differentially expressed target genes which included gene promoter methylation status, TF down- or upregulation, targeting of hyper or hypomethylated genes only or both and individual TF in the model found highly significant association between promoter methylation status and gene expression (p value <0.05), as expected based on previous chi-square results. It also detected borderline significant association between expression level and up-regulation of TF (p value=0.05) and a significant association between expression levels and targeting by FOXF2 (p value <0.0001), PITX2 (p value =0.01), RUNX1 (p value=0.01) and SMAD3 (p value =0.044).

6.3.14. Comparison of genome-wide methylation of normal skin and cSCC reveals 3512 hypermethylated and 1625 hypomethylated probes

The cut-off for differentially methylated probes comparing normal skin and cSCC was set at adjusted p-value <0.0001 and revealed a total of 5137 DMPs. A total of 18 cSCC samples and 20 normal skin samples were included in this analysis, and 4 tumours and normal skin samples originated in the same patient. Of 5137 DMPs, 3512 (68%) were hypermethylated in cSCC, while

1625 (32%) were hypomethylated in cSCC (Figure 6.26). Heatmap of beta-values corresponding to this probe set are shown Figure 6.27.

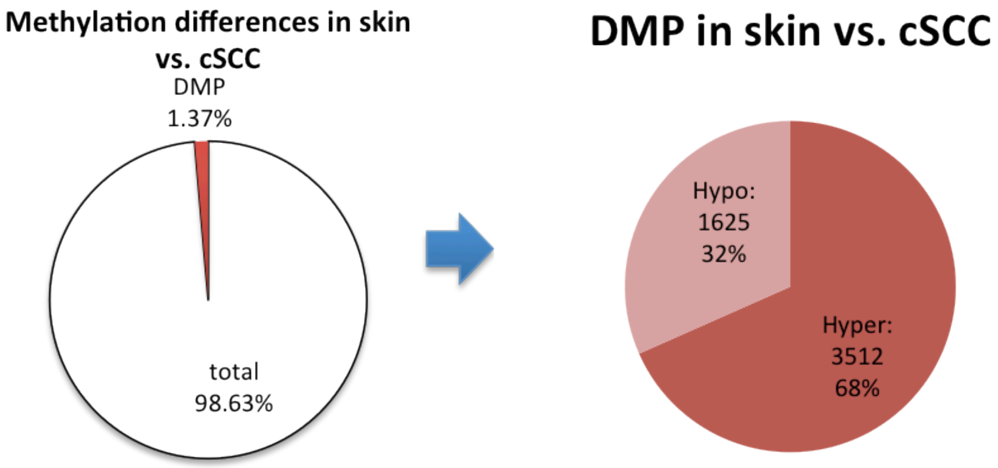


Figure 6.26 Percentage of differentially methylated probes (DMP) in cSCC compared to skin. DMP=differentially methylated probes. Of DMPs, 68% are hypermethylated in cSCC, 32% are hypomethylated compared to normal skin.

.

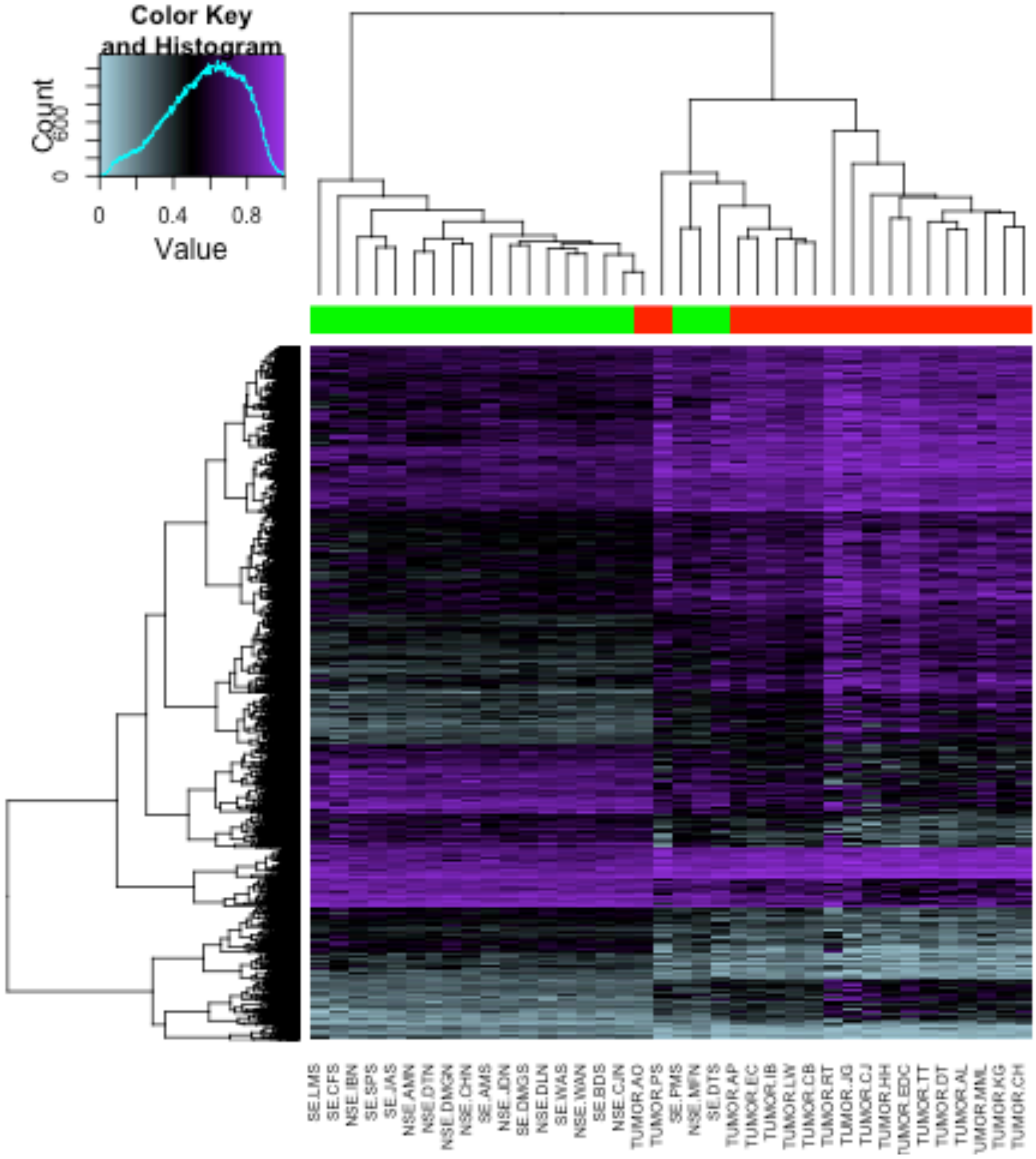


Figure 6.27 Heatmap of differentially methylated probes in cSCC. Dendrograms are based on unsupervised clustering of beta-values across all cSCC and normal skin samples. With the exception of 2 cSCC and three skin samples, this probeset was able to distinguish cSCC from normal skin. Red samples=cSCC, green samples=normal skin.

The hypermethylated probes translated into 1812 unique genes, while hypomethylated probes aligned to 819 unique genes. 126 genes appeared in both lists due to the presence of multiple probes corresponding to various regions of the gene on the array. Gene set enrichment analysis of hyper- and hypomethylated genes was then conducted separately.

6.3.15. Functional genomic distribution of differentially methylated probes

Functional genomic distribution of hypermethylated and hypomethylated probes (Figure 6.28, Table 6.14) shows that hypermethylated probes correspond predominantly to gene bodies, while 27% of hypomethylated probes corresponded to intergenic regions.

Skin vs. cSCC	Hypo	Hyper	P value
intergenic	644	219	<0.0001
Promoter	582	1594	0.001
3'UTR	62	509	<0.0001
Body	1109	3401	<0.0001

Table 6.14 Functional annotation of hyper- and hypomethylated probes in cSCC. P -value based on Chi-square test with Yates' correction.

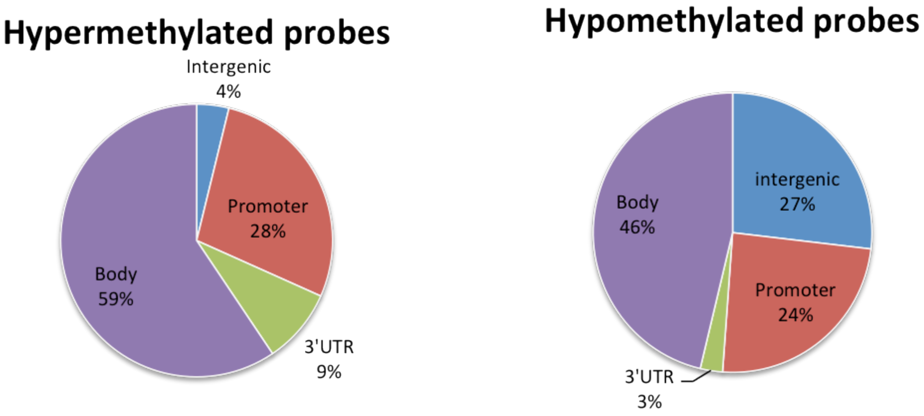
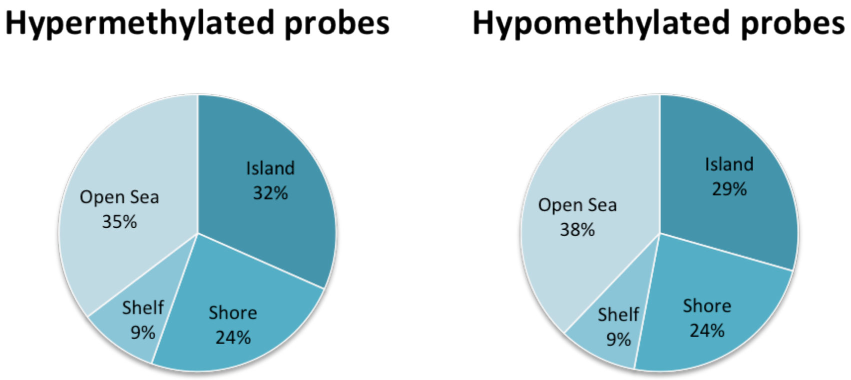


Figure 6.28 Functional genomic distribution of hyper- and hypo-methylated genes. Hypomethylation of probes corresponding to intergenic regions is far more common compared with hypermethylation.

6.3.16. CpG islands and the neighbourhood context analysis shows similar distribution of both hyper- and hypomethylated probes

The distribution of probes with respect to their positional relationship to CpG islands has revealed no differences in the proportion of probes corresponding to Shelf and Shore, and three percentage points increase in probes corresponding to CpG islands among hypermethylated probes and three percentage points decrease in Open Sea probes among this set of probes.



6.3.17. Pathway analysis of hyper- and hypomethylated methylated genes shows a substantial overlap with pathways dysregulated by hyper- and hypomethylated genes in AK

Pathway analysis of hypermethylated genes revealed a total of 84 dysregulated pathways with p value <0.05. A full list is provided in Appendix 15. Seventy (83%) of these overlapped with pathways dysregulated through hypermethylated probes in AK. Figure 6.29 displays pathways with 20 or more annotated genes.

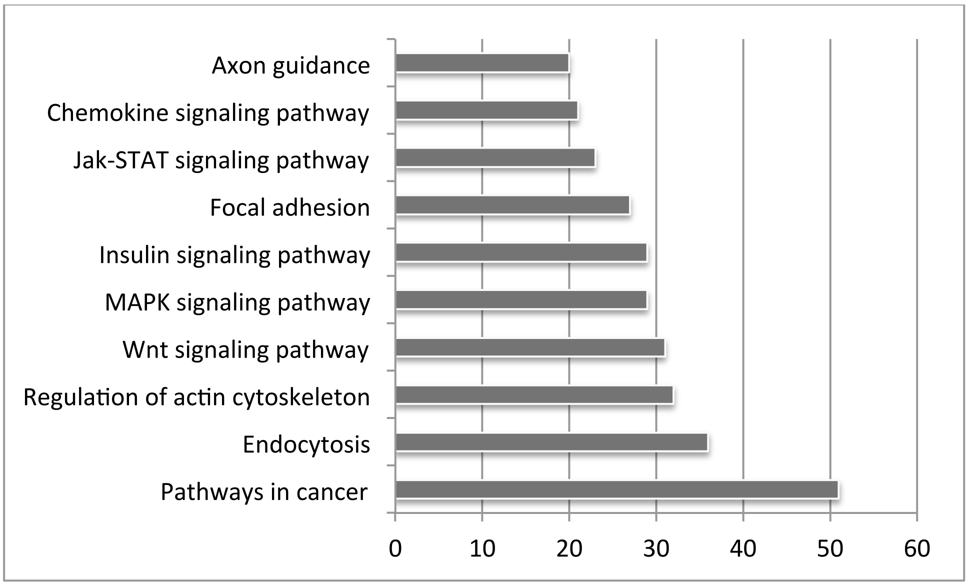


Figure 6.29 Dysregulated KEGG pathways in genes hypermethylated in cSCC with 20 or more annotated genes.

Table 6.15 lists selected dysregulated pathways in hypermethylated genes that are directly relevant in oncogenesis.

Description	**# Genes**	**Adjusted p value**
Pathways in cancer	51	5.31E-15
Wnt signaling pathway	31	1.34E-12

MAPK signaling pathway	29	5.00E-06
Focal adhesion	27	2.69E-07
Jak-STAT signaling pathway	23	4.25E-07
Tight junction	19	7.27E-06
ErbB signaling pathway	18	8.73E-08
Chronic myeloid leukemia	17	4.67E-08
Prostate cancer	17	5.58E-07
PPAR signaling pathway	16	9.88E-08
mTOR signaling pathway	15	1.87E-08
Pancreatic cancer	15	6.07E-07
Cell cycle	15	5.26E-04
Acute myeloid leukemia	14	5.63E-07
Colorectal cancer	14	7.86E-07
Small cell lung cancer	14	2.46E-05
TGF-beta signaling pathway	14	3.06E-05
Apoptosis	13	1.60E-04
Non-small cell lung cancer	12	6.04E-06
Epithelial cell signaling in Helicobacter pylori infection	12	5.56E-05
Adherens junction	12	1.47E-04
VEGF signaling pathway	12	1.60E-04
Notch signaling pathway	9	3.01E-04
Basal cell carcinoma	9	8.13E-04
Hedgehog signaling pathway	9	9.03E-04
p53 signaling pathway	7	4.13E-02

Table 6.15 Selected dysregulated pathways related to cancer in genes hypermethylated in cSCC.

KEGG pathway analysis of hypomethylated genes revealed 52 dysregulated pathways with p value <0.05 (listed in Appendix 16). 45 (87%) of those overlapped with pathways dysregulated in genes hypomethylated in AK, and 44 (85%) of those also overlapped with pathways dysregulated in hypermethylated genes in cSCC. Table 6.16 lists pathways that overlap between hyper- and hypomethylated genes, and those that are uniquely dysregulated in either group. Figure 6.30 lists pathways dysregulated in hypomethylated genes with 15 or more annotated genes.

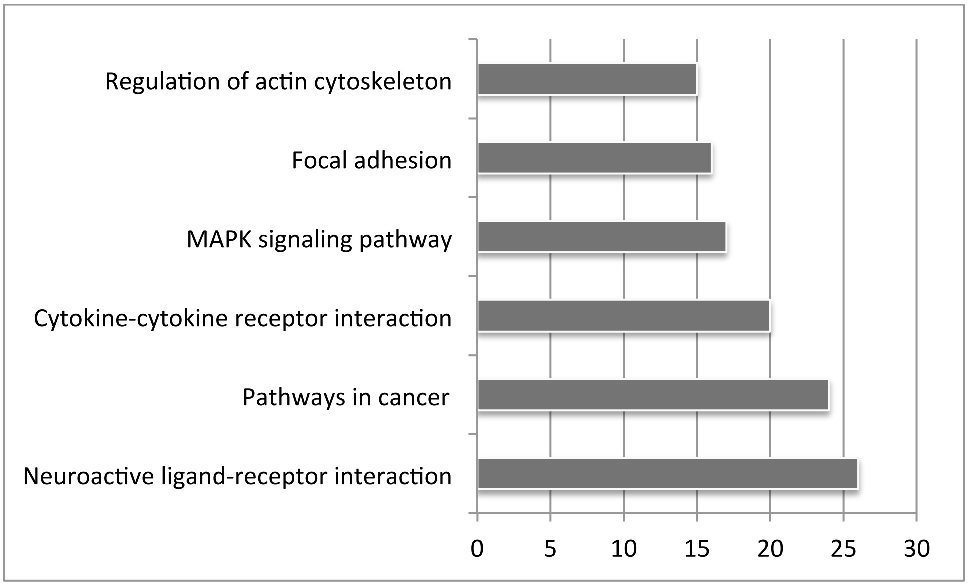


Figure 6.30 Dysregulated pathways in hypomethylated genes in cSCC with 15 or more annotated genes.

Shared	Hyper only	Hypo only
Adipocytokine signaling pathway	Acute myeloid leukemia	Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Aldosterone-regulated sodium reabsorption	Adherens junction	Cardiac muscle contraction
Axon guidance	Amino sugar and nucleotide sugar metabolism	Cell adhesion molecules (CAMs)
Basal cell carcinoma	Aminoacyl-tRNA biosynthesis	Glycolysis / Gluconeogenesis
Calcium signaling pathway	Apoptosis	Glycosphingolipid biosynthesis - lacto and neolacto series
Chemokine signaling pathway	Biosynthesis of unsaturated fatty acids	Glyoxylate and dicarboxylate metabolism
Cytokine-cytokine receptor interaction	Bladder cancer	Histidine metabolism
Dilated cardiomyopathy	Cell cycle	Neuroactive ligand-receptor interaction
ECM-receptor interaction	Chronic myeloid leukemia	
Endocytosis	Circadian rhythm - mammal	
Endometrial cancer	Colorectal cancer	
ErbB signaling pathway	Epithelial cell signaling in Helicobacter pylori infection	
Fc epsilon RI signaling pathway	Ether lipid metabolism	
Fc gamma R-mediated phagocytosis	Fatty acid metabolism	
Focal adhesion	Glioma	
Fructose and mannose metabolism	Glycerolipid metabolism	
Gap junction	Glycerophospholipid metabolism	
GnRH signaling pathway	Glycosaminoglycan biosynthesis - chondroitin sulfate	

Hedgehog signaling pathway	Glycosaminoglycan biosynthesis - heparan sulfate
Hypertrophic cardiomyopathy (HCM)	Glycosaminoglycan biosynthesis - keratan sulfate
Inositol phosphate metabolism	Hematopoietic cell lineage
Insulin signaling pathway	Leukocyte transendothelial migration
Jak-STAT signaling pathway	Lysine degradation
Long-term depression	Lysosome
Long-term potentiation	Non-small cell lung cancer
MAPK signaling pathway	Notch signaling pathway
Melanogenesis	Oocyte meiosis
Melanoma	PPAR signaling pathway
NOD-like receptor signaling pathway	Phosphatidylinositol signaling system
Neurotrophin signaling pathway	Primary bile acid biosynthesis
Pancreatic cancer	Prostate cancer
Pathways in cancer	Purine metabolism
Progesterone-mediated oocyte maturation	Pyruvate metabolism
Regulation of actin cytoskeleton	Small cell lung cancer
Renal cell carcinoma	Steroid biosynthesis
TGF-beta signaling pathway	T cell receptor signaling pathway
Thyroid cancer	VEGF signaling pathway
Tight junction	Vasopressin-regulated water reabsorption
Toll-like receptor signaling pathway	Vibrio cholerae infection
Type II diabetes mellitus	p53 signaling pathway
Ubiquitin mediated proteolysis	
Vascular smooth muscle contraction	
Wnt signaling pathway	
mTOR signaling pathway	

Table 6.16 Dysregulated KEGG pathways in both hyper- and hypomethylated genes, with overlapping pathways listed in the first column.

6.3.18. Positional gene signature of hyper- and hypomethylated genes reveals overrepresentation of DMP on chromosomes 1, 3, 16 and 17

Positional gene signature of hypo- and hypermethylated genes presented in Table 6.17 shows overlapping regions on chromosomes 1,3,7 and 8, and a significant overrepresentation of hypermethylated genes in multiple bands on chromosomes 3, 11, 12, 16, and 17.

Shared	hyper	hypo

chr1p36	chr10q26	chr10q21
chr1q32	chr11p11	chr11p15
chr3q21	chr11q13	chr13q32
chr7p22	chr12q13	chr20q13
chr8q24	chr12q24	chr21q22
	chr15q22	chr2p25
	chr16p13	chr6q25
	chr16q13	chr7p12
	chr16q22	
	chr16q24	
	chr17p11	
	chr17p13	
	chr17q21	
	chr17q25	
	chr19p13	
	chr1p34	
	chr1q21	
	chr1q22	
	chr22q13	
	chr2p23	
	chr3p21	
	chr3p25	
	chr4p16	
	chr5q35	
	chr6p21	
	chr8p21	
	chr9q34	

Table 6.17 Positional gene signature of differentially methylated genes in cSCC.

I then integrated overlapping regions with regions of known genomic changes in cSCC (described in greater details in Chapter 4) in order to find regions in which epigenetic changes detected by the array may be due to genomic imbalance (Table 6.18).

Chromosome	Location	Positional signature	Methylation status
Chr1			
Type			
G	1pter-1p32.3	1p36; 1p34	both; hyper
G	1qcen-1qter	1q32; 1q21; 1q22	both; hyper; hyper
Chr3			
G	3qcen-3qter	3q21	both
L	3pter-3pcen	3p21; 3p25	hyper; hyper
Chr4			
L	4pter-4p13	4p16	hyper

Chr5			
L	5qcen-5qter	5q35	hyper
Chr6			
G	6pter-6p21.1	6p21	hyper
Chr7			
G	7pter-7pcen	7p22; 7p12	both; hypo
Chr8			
L	8pter-8pcen	8p21	hyper
G	8qcen-8qter	8q24	both
Chr9			
G	9qcen-9qter	9q34	hyper
Chr10			
L	10pter-10qter	10q26; 10q21	hyper; hypo
Chr11			
L	11pter-11pcen	11p11; 11p15	hyper; hypo
G	11q12.2-11q13.4	11q13	hyper
Chr13			
L	13qcen-13qter	13q32	hypo
Chr16			
L	16p13.11-16pcen	16p13	hyper
Chr17			
L	17pter-17qter	17p11; 17p13	hyper; hyper
Chr19			
L	19pter-19p13.2	19p13	hyper
G	19pter-19pcen	19p13	hyper
Chr20			
G	20pter-20qter	20q13	hypo
Chr22			
L	22q11.23-22qter	22q13	hyper
G	22qcen-22qter	22q13	hyper

Table 6.18 Integration of positional gene signature in cSCC with segments of genomic imbalance in cSCC. L=loss of genomic material, G=gain of genomic material.

6.3.19. Overlap between hypermethylated and hypomethylated DMP in AK and cSCC is 34.5% and 49.3%, respectively, with functional genomic distribution similar to DMP in cSCC

As expected, a substantial overlap was detected between DMP in AK and SCC: 802/1625 (49.3%) of hypomethylated and 1213/3512 (34.5%) of hypermethylated probes. The overlapping hyper- and hypomethylated probes translated into 751 and 433 unique genes, respectively.

As shown in Table 6.19 and Figure 6.31, functional genomic distribution (FGD) of overlapping DMPs was largely similar to FGD of probes dysregulated in cSCC.

Overlapping probes	hyper	hypo	P value
Intergenic	214	318	<0.0001
Promoter	670	309	0.0008
3'UTR	158	36	<0.0001
Body	1123	551	0.0003

Table 6.19 Functional genomic distribution of DMP overlapping between skin vs. AK and skin vs. cSCC.

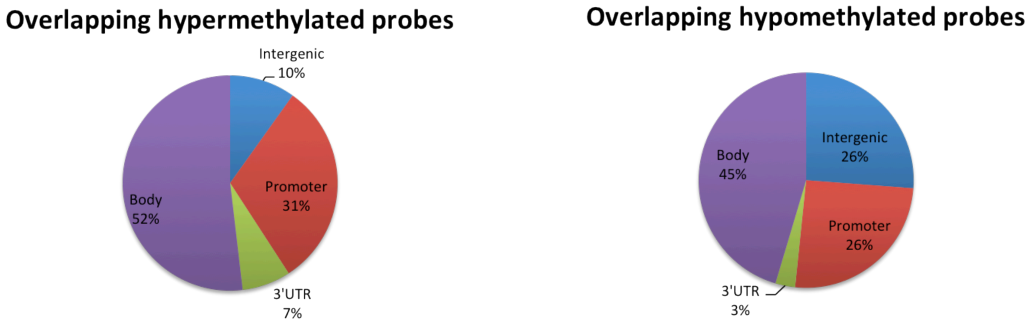


Figure 6.31 Functional genomic distribution of hyper- and hypomethylated probes overlapping in AK and cSCC. As in probes dysregulated in cSCC, hypomethylated probes more commonly correspond to intergenic regions, while 3'UTR and body probes are more commonly hypermethylated.

6.3.20. CpG Island and the neighbourhood context of overlapping DMPs reveals the predominance of Open Sea probes

In a stark contrast to both DMPs in AK and cSCC, overlapping probes corresponded mainly to Open Sea regions (68% of all overlapping hypomethylated probes), and minimally to CpG islands, which represented approximately 30% of DMPs in cSCC.

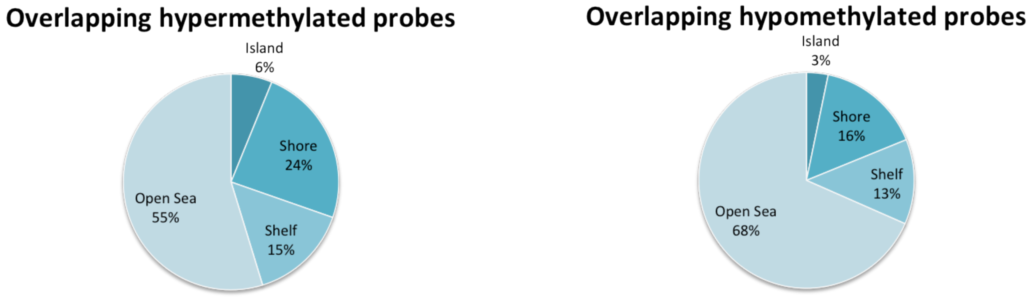


Figure 6.32 Overlapping probes corresponding to CpG islands and neighbourhood regions.

6.3.21. Comparison of overlapping hyper- and hypomethylated genes to published signatures of chemical and genetic perturbations shows significant overlaps with genes regulated by critical oncogenic molecules

Comparison of overlapping DMPs to curated chemical and genetic perturbations in GSEA based on published data revealed 54 hypomethylated genes to overlap with genes upregulated in nasopharyngeal carcinoma (Dodd et al., 2006), 44 to correspond to genes upregulated by MAPK8 (Yoshimura et al., 2005), 36 to be targets of SUZ12 and 21 to be targeted by PRC2 based on ChIP on chip data in hESC (Ben-Porath et al., 2008).

Additional GSEA analysis of curated chemical and genetic perturbations also revealed that eighty hypermethylated genes overlapped with TP53 targets in mammary epithelium (Perez et al., 2007) and 46 were shown to be downregulated in mouse skin upon RB1 and TP53 knock-down (Martinez-Cruz et al., 2008). Additionally, 38 genes were shown to be SOX2 targets in hESC (Ben-Porath et al., 2008) and 37 were shown to have their promoter bound by FOXP3 based on ChIP-chip analysis (Zheng et al., 2007). These detections were based on significant overlaps with curated datasets included in the MsigDB database.

6.3.22. Integration with expression data based on methylation differences reveals decreased gene expression with increased methylation

As shown in Figure 6.33, similar to AK, increase in methylation leads to decreased gene expression, but this trend changes around the extreme values towards increased expression even in hypermethylated genes.

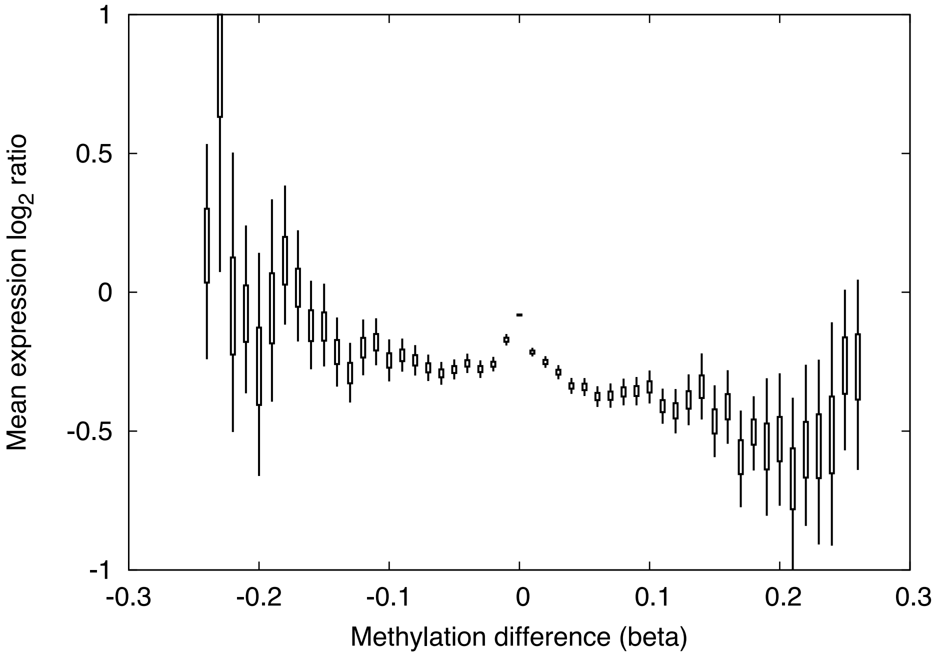


Figure 6.33 Mean expression log2 ratio in cSCC correlated with different levels of methylation difference compared to normal skin. Negative values=hypomethylated in cSCC. As observed in

AK, with increased methylation, expression levels of genes generally decrease, with the exception of extreme values (differences greater than 20%) which lead to universal expression increase irrespective of hyper- of hypomethylation. Peak observed around 0 is most likely an artifact due to the very low number of genes with minimal or no differences in methylation. Figure prepared in collaboration with Dr Thomas Down.

6.3.23. Validation of 450K methylation data

DMP detected with 450K array were validated using 10 laser-capture microdissected cSCC samples and paired normal skin originating in male RTR patients. As described in Chapter 2, methylation levels of regions within 50 bp of DMP were averaged and subtracted to obtain differences in methylation.

In case of perfect validation, Figure 6.35 would replicate Figure 6.34. Although Figure 6.35 shows a substantial overlap between hyper- and hypomethylated probes and centers around zero, important differences are maintained in both hyper- and hypomethylated probes and these probes remain well-segregated in areas of higher methylated differences.

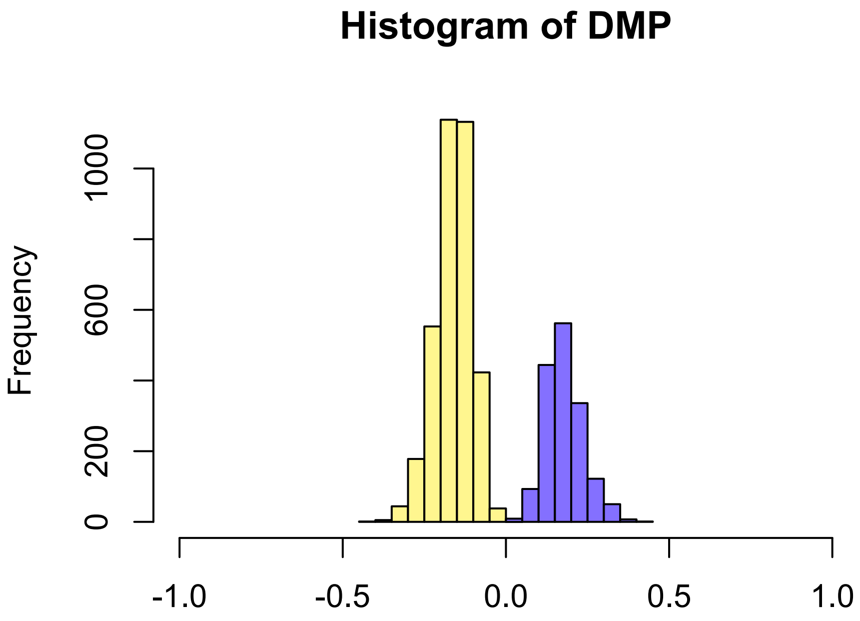


Figure 6.34 Histogram of differences in beta-values in DMPs in cSCC based on Illumina 450K data. Hypomethylated probes are shown in blue, hypermethylated are shown in yellow. X axis=differences in methylation.

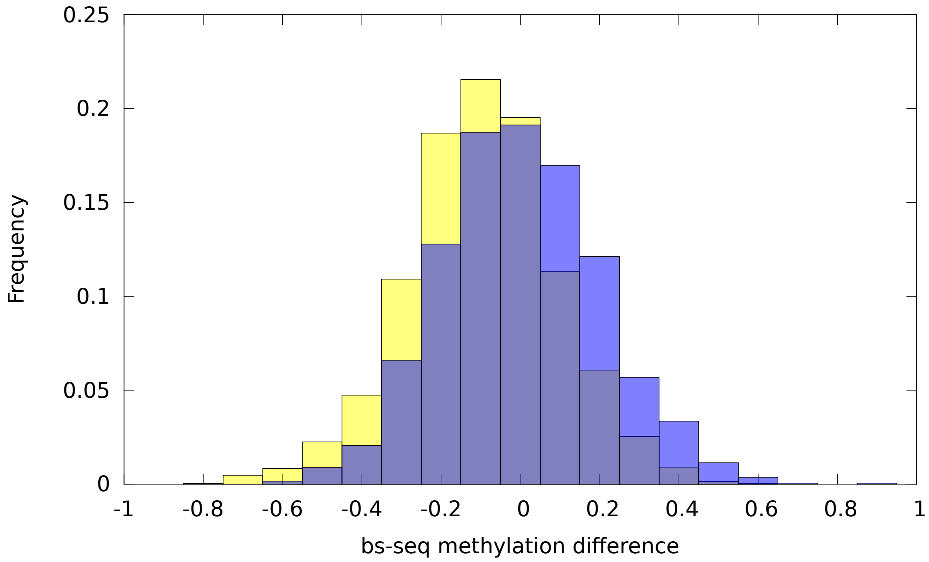


Figure 6.35 Histograms of bs-seq methylation differences corresponding to DMP in cSCC. Hypomethylated regions are shown in blue, hypermethylated are shown in yellow, and overlapping areas are shown in purple. Figure prepared in collaboration with Dr Thomas Down.

6.3.24. Comparison of DMP in cSCC with cSCC cell line methylation array data shows 21.3% overlap and absolute consistency in methylation direction in overlapping probes

Overlap between 398 DMP in cSCC cell lines (described in greater details in Chapter 5) and DMP in cSCC clinical samples found 85 (21.3%) overlapping probes listed in Table 6.20. All probes hyper- or hypomethylated in cSCC cell lines were consistently hyper- or hypomethylted in cSCC clinical samples.

Overlapping DMP	**Methylation difference (cSCC)**	**Methylation difference (cSCC cell lines)**	**Gene**
cg19884658	-0.275119943	-0.806527778	**'KLHL21'**
cg00283535	-0.25849724	-0.385046296	**'ANXA13'**
cg06615154	-0.234064056	-0.420555556	**'S100A3'**
cg17071957	-0.223950751	-0.454351852	'GSN'
cg19664945	-0.211805094	-0.6425	'GPR75'
cg02774160	-0.208095751	-0.414398148	'GGT1'
cg14423778	-0.207495748	-0.622777778	**'MBNL1'**
cg10725344	-0.191535391	-0.550740741	'UNQ1940'
cg19881895	-0.186313042	-0.668148148	'SLC43A3'
cg17777592	-0.179162659	-0.511388889	**'CAB39L'**
cg10894512	-0.155323456	-0.506851852	**'ACTA2'**
cg12111714	-0.149834427	-0.627592593	**'ATP8A2'**
cg11928198	-0.148705114	-0.520185185	'SCNM1'
cg25545088	-0.146568608	-0.355092593	'ABCC10'
cg00891541	-0.146318331	-0.483425926	**'SMPD3'**
cg06656924	-0.144614617	-0.480694444	**'ZNF76'**

cg02164046	-0.143679321	-0.69875	'SST'
cg20585500	-0.139206272	-0.119212963	'GPHA2'
cg15901783	-0.139030221	-0.53287037	'KCTD12'
cg16708623	-0.13734916	-0.566296296	'TRIM2'
cg10261191	-0.135532024	-0.715092593	'MGC39545'
cg26185508	-0.133044738	-0.505324074	'CDCP2'
cg24475171	-0.132128841	-0.546712963	**'C9orf78'**
cg17568996	-0.131679975	-0.795648148	'NFAM1'
cg19297823	-0.131227901	-0.432546296	'GCNT4'
cg25400358	-0.125514123	-0.338657407	'GPR137'
cg13678049	-0.124327379	-0.576574074	'PARC'
cg26523005	-0.119369449	-0.30587963	'ZNF662'
cg09106999	-0.117949134	-0.378240741	**'CDK2'**
cg18267381	-0.117660114	-0.55212963	'ZNF659'
cg26583078	-0.117641564	-0.453148148	**'SORBS2'**
cg22752533	-0.115850658	-0.546574074	**'SLC12A5'**
cg06092815	-0.115838579	-0.636064815	'SKIP'
cg17496921	-0.114575113	-0.598472222	**'TSPAN16'**
cg03289872	-0.11386	-0.622175926	'ZNF667'
cg21342728	-0.109232916	-0.774259259	**'GPR24'**
cg24576425	-0.107214168	-0.588425926	'GALNT5'
cg14642338	-0.105964447	-0.101851852	'DKFZP586H2123'
cg03355526	-0.105609264	-0.515648148	'ZNF454'
cg21697134	-0.10513978	-0.758888889	**'FN3K'**
cg04008913	-0.101546957	-0.431666667	'SNRK'
cg02647265	-0.100980506	-0.567222222	'FLJ22471'
cg23713520	-0.099507042	-0.654027778	**'TM4SF11'**
cg19713460	-0.094102444	-0.642175926	**'SYNGR1'**
cg13288195	-0.090652605	-0.697222222	'FBXL22'
cg26104206	-0.087918025	-0.47787037	'PHYHIP'
cg00594952	-0.08606512	-0.678472222	'RIMS3'
cg22918700	-0.085485741	-0.208472222	**'LOC161247'**
cg22681784	-0.077257113	-0.534166667	**'SPINK2'**
cg10559803	-0.076638426	-0.278333333	'RALGPS2'
cg23807646	-0.075254238	-0.448287037	**'SLC26A8'**
cg13870866	-0.072867702	-0.6025	'TBX20'
cg18254848	0.060475359	0.525462963	'CLC'
cg11827101	0.075420942	0.58712963	'LOC339789'
cg14086122	0.084213629	0.446435185	'MGC7036'
cg02110963	0.086314257	0.455925926	'MGC40222'
cg11091262	0.095910565	0.468611111	'GALNTL5'
cg10054857	0.097116639	0.541388889	**'C18orf20'**
cg05208878	0.098682565	0.449351852	'PRAME'
cg14757492	0.101085904	0.711666667	'DDX49'
cg27168844	0.101673966	0.470092593	**'IL17'**

cg24468890	0.103874202	0.233333333	'HDAC1'
cg07816074	0.10594635	0.864259259	'SH3TC1'
cg23109897	0.107002313	0.171712963	'PRKDC'
cg18740800	0.11015179	0.369953704	**'KRT7'**
cg08675585	0.110203353	0.65287037	**'C14orf166B'**
cg08901867	0.110438087	0.43962963	'TTLL6'
cg08886154	0.114591013	0.467222222	**'PAX4'**
cg22988566	0.117335619	0.593009259	**'WFDC10B'**
cg20950277	0.11793971	0.623935185	'TNIP3'
cg13435792	0.122004147	0.456805556	'C12orf46'
cg06793062	0.126807469	0.508333333	**'CNTNAP4'**
cg00436282	0.127118992	0.713055556	'STATH'
cg05429895	0.129041385	0.523518519	'TLR4'
cg00958560	0.143420433	0.556851852	**'FLJ46481'**
cg23254045	0.144069478	0.471574074	**'TMCO5'**
cg20287234	0.147602945	0.675555556	'GPR55'
cg16001913	0.148394503	0.466018519	'HK1'
cg24423088	0.15170408	0.641574074	'KRTAP8
cg21032583	0.159286071	0.665324074	'LMLN'
cg26757793	0.161896045	0.514768519	**'MMP20'**
cg16812893	0.178286492	0.464259259	'KRTAP15
cg02833180	0.181942487	0.617777778	**'PLCL1'**
cg26799474	0.206382182	0.836203704	'CASP8'
cg06051311	0.231622336	0.662592593	'TRIM15'

Table 6.20 DMP overlapping between cSCC clinical samples and cSCC cell lines. Negative value indicates hypermethylation in comparison to normal skin or normal cultured keratinocytes. Genes highlighted in bold were validated with bs-sequencing.

Using bs-seq data to validate this set of genes, I obtained methylation levels for 74 genes, of which 32 (43%) showed hyper- or hypomethylation consistent with the microarray results. Validated genes are highlighted in bold in Table 6.20.

6.3.25. Comparison of AK and cSCC reveals 81 hypermethylated and 23 hypomethylated genes in cSCC

Comparison of cSCC and AK data detected a total of 148 DMP in cSCC, of which 108 were hyper- and 40 hypomethylated in cSCC. Full list of DMP with corresponding genes is provided in Appendix 17. Figure 6.36 shows heatmap of this probeset.

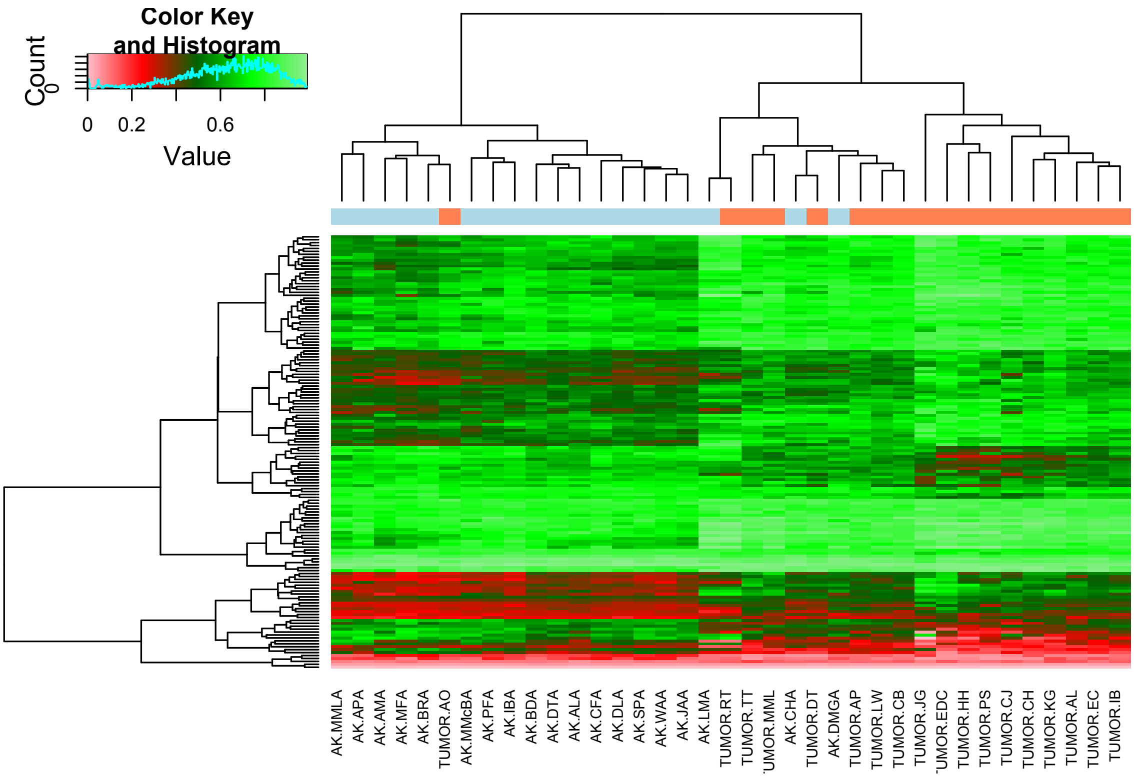


Figure 6.36 Heatmap of probes differentially methylated in cSCC compared to AK. Light blue=AK, coral=cSCC. This set of 148 probes leads to almost complete separation of AK and cSCC, with the exception of 1 cSCC and 3 AK.

GO term analysis of hypermethylated genes revealed over-representation of genes involved in the synthesis of RNA on a template of DNA and other processes related to gene transcription regulation. The most important genes involved in gene transcription included SREBF1, ILF3, EHF, MED26, SMARCD2, NOTCH4, LIMD1, NFIX, NFIC, and MED24. Additional GSEA analysis detected 11 SMAD2 or SMAD3 targets (ACADVL, BLCAP,CASZ1, EHF, SMAD6, KIAA1737, KIAA0182, SREBF1, PPP2R2D, SSBP3, BCL9L) based on published data (Koinuma et al., 2009), TP63 targets (ZNF385A, ASS1, CASZ1, SMAD6, SEMA6D, CRAMP1L, KLHL21) and TP53 targets (BLCAP, ALDH4A1, ZNF385A, SMAD6, KIAA1737, KIAA0182, FAM53B, CRAMP1L, KLHL21, LIMD1, TOLLIP) many of which contain DNA binding sites for both TF (Perez et al., 2007), and 5 genes down-regulated in invasive breast cancer: ERBB3, SEC61A1, PPP2R2D, ILF3, AP2A2 (Wang et al., 2007).

Analysis of hypomethylated genes detected four genes upregulated in Ewing's sarcoma: ADARB1, SV2B, PCSK2, PBX1 (Riggi et al., 2008) and 2 genes that confer resistance to trabectedin, an anti-tumour drug, in soft tissue tumours MAML3 and DUT (Marchini et al., 2005).

6.3.26. Functional genomic distribution of DMP in cSCC compared to AK shows hypomethylation of intergenic regions and more frequent body and 3'UTR hypermethylation

Similar to DMPs in cSCC versus normal skin comparison, most hypermethylated probes corresponded to gene bodies, and 3'UTR probes were also more commonly hypermethylated. Intergenic probes were frequently hypomethylated, and no hypomethylated 3'UTR probe was detected.

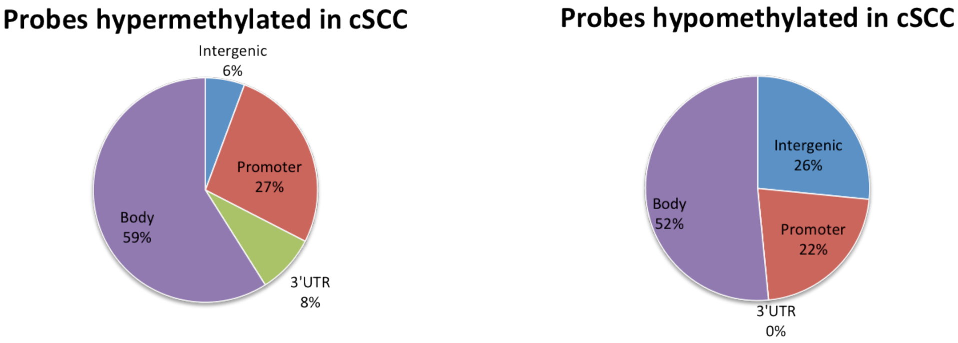


Figure 6.37 Functional genomic distribution of DMP in cSCC compared to AK. Similar to probes dysregulated in cSCC compared to skin, hypomethylated probes more commonly correspond to intergenic regions and no 3'UTR probe was detected as hypomethylated, while 3'UTR and body probes are more commonly hypermethylated.

6.3.27. CpG islands and the neighbourhood context analysis of DMP in cSCC compared to AK shows few probes corresponding to CpG islands, and predominance of Open Sea probes

As shown in Figure 6.38, probes corresponding to CpG islands were the least prevalent among DMPs, while Open Sea probes were the most common, representing 63% of all hypomethylated probes. One third of hypermethylated probes corresponded to Shore regions.

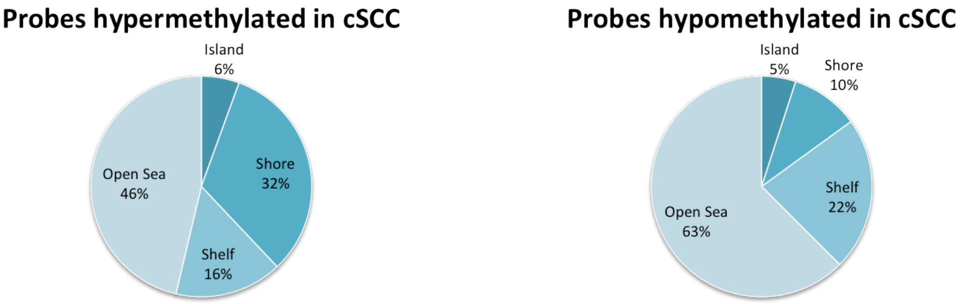


Figure 6.38 CpG islands and neighbourhood regions analysis of DMP in cSCC. Both hyper- and hypomethylated probes show low representation of CpG islands probes. Almost a third of hypermethylated probes corresponded to Shore regions, and the majority of hypomethylated probes corresponded to Open Sea regions.

6.3.28. Comparison of DMGs with differentially expressed genes reveals only a modest overlap

Only 22 genes (20%) overlapped with DEGs (AK vs. cSCC comparison). Table 6.21 lists overlapping genes with methylation status in cSCC and expression fold-change in cSCC.

Overlaping genes	Methylation in cSCC	Expression in cSCC (FC)
ACADVL	Hyper	-0.59490742
ASS1	Hyper	-0.97591401
C1orf133	Hypo	-0.649103
CASZ1	Hyper	-1.067690085
ERBB3	Hyper	-0.812783857
EXPH5	Hyper	-1.137355733
ILF3	Hyper	0.4488273
KIAA0182	Hyper	-0.530262499
MAML3	Hypo	-0.824805407
MCM7	Hyper	0.769505193
MRVI1	Hyper	-0.867038372
MTHFD2	Hypo	0.497488708
NFIX	Hyper	-0.966907945
PBX1	Hypo	-0.944261999
PCSK2	Hypo	-1.784293921
PTBP1	Hyper	0.654671365
SBNO2	Hyper	0.824146874
SEC61A1	Hyper	0.711746567
SMAD6	Hyper	-0.773196057
TPRG1	Hyper	-1.172766618
VPS37B	Hyper	0.55773373
ZNRF1	Hyper	-0.976904048

Table 6.21 Overlap between differentially methylated and differentially expressed genes in cSCC compared to AK. Negative foldchange represents down-regulation in cSCC compared to AK.

6.3.29. Methylation profile of cSCC samples is not associated with cSCC subtyping nor with immune status

The dataset included 9 WD, 1 WMD, 5 MD and 2 PD cSCCs (1 cSCC was of unknown histology and was excluded from this analysis). Neither clustering analysis (Figure 6.39) nor association calculation using the pvclust and CpGAssoc R packages revealed significant association with cSCC histological typing; either with the above-described histological phenotype description or combining all non-WD tumours into “less than WD” group.

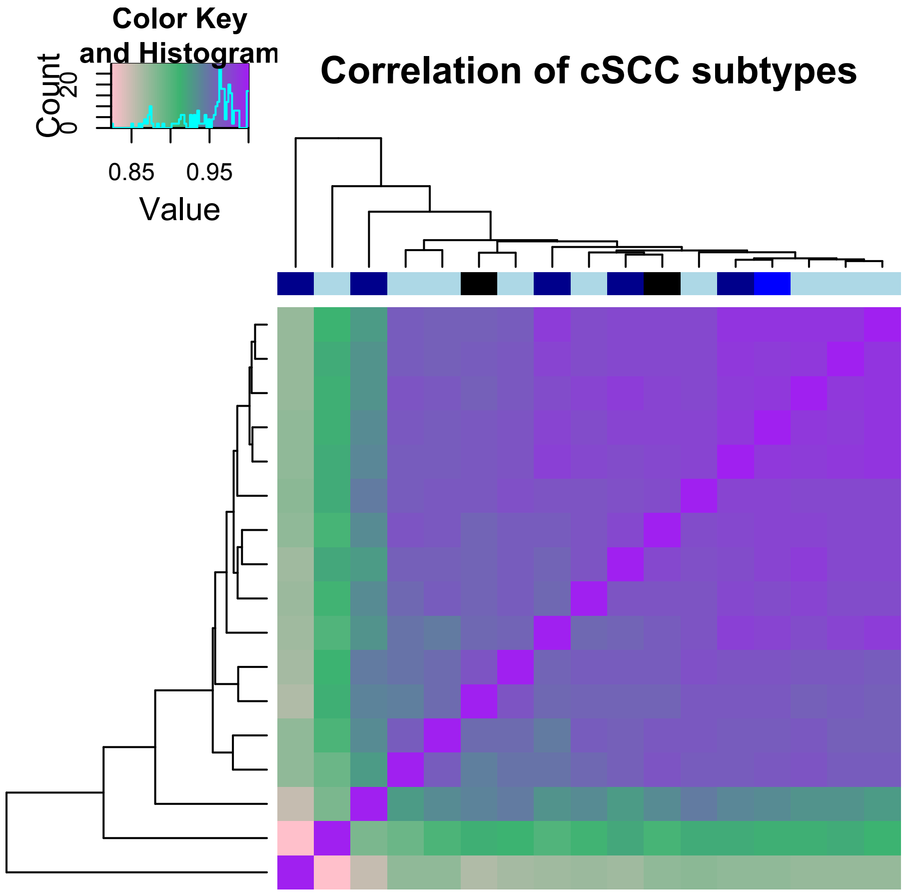


Figure 6.39 Correlation of different cSCC subtypes. Light blue=WD, blue=WMD, dark blue=MD, black=PD. No distinct clustering pattern of cSCC samples is apparent. The left-most MD cSCC originated in PUVA-treated patient.

CpGAssoc package detected no association between the immune status nor gender and cSCC methylation profile, but it detected an association with phenotype cSCC originating in a PUVA patient. This indicates that statistically significant difference in the methylation of this tumour compared to the remaining samples.

6.4. Discussion

This is the first study of genome-wide methylation in normal skin, actinic keratosis and cSCC using state-of-the-art Illumina 450K methylation array. Additionally, we generated the first whole-genome bisulfite sequencing data of laser-capture micro-dissected cSCC and paired normal skin.

6.4.1. Data processing and normalisation

Given that the new Illumina 450K array contains two types of probes, it has been previously recommended to normalise the two probe types separately (Pidsley et al., 2013), in addition to alternative computational approaches such as Subset-quantile Within Array Normalisation (Maksimovic et al., 2012). The purpose of the study was to detect the most pronounced differences in methylation between normal and precancerous and cancerous tissues, which are maintained even after aggressive normalisation methods applied to the 450K methylation data (Morris and Lowe, 2012), thereby quantile normalisation was chosen as a consensus method previously applied to 450K methylation data derived from cancer samples (Cahill et al., 2013).

6.4.2. The absence of differences in methylation between sun-exposed and non-sunexposed skin is concordant with the previous literature

I found no statistically significant differences in methylation between NSE and SE samples. This is consistent with previous reports that showed no methylation difference in cultured keratinocytes following UV radiation (Lahtz et al., 2013). Although UV light is a known mutagen, it seems likely that the vast majority of UV-induced changes in methylation are either immediately repaired, or methylation is largely resistant to the effects of UV light, perhaps as an evolutionary advantage. This may also indicate that the vast differences in methylation observed in cSCC samples compared to normal skin are possibly not due to the direct effect of UV light on methylation regulatory patterns.

In contrast, comparison of the transcriptional profile of NSE and SE has revealed differences between the two tissues (details are described in Chapter 3). Probes for all differentially transcribed genes are present on the methylation array. The lack of differences detected in methylation profiles of SE and NSE skin may be due to differences in half-life of mRNA and the stability of methylation changes. They may equally reflect alternative epigenetic regulatory mechanisms, such as micro RNA, in response to UVR. This lack of concordance may also be due to technical issues, since the transcriptional profile was derived from samples that were treated with RNALater, or due to data processing since differences in methylation were detected with non-parametric Welch t-test, while transcriptional differences were determined with Bayesian statistics.

6.4.3. Functional genomic distribution of differentially methylated probes in AK reveals more common hypermethylation in gene body, promoter and 3'UTR regions, and hypomethylation in intergenic regions

Our results indicate that hypermethylated probes correspond more commonly to gene body, promoter and 3'-UTR regions. Hypermethylation of gene promoters is commonly observed in cancer (Sproul et al., 2012). Hypomethylation of gene bodies and 3'-UTR regions correlates with decreased gene expression in breast cancer (Shann et al., 2008), and widespread gene body hypomethylation was previously observed in chronic lymphocytic leukaemia using the 450K array (Kulis et al., 2012). In our data, we see similar trend for probes corresponding to 3'UTR, in which hypomethylation leads to decreased expression in AK, while hypomethylation of gene bodies and promoters is associated with increased gene expression.

Most pronounced differences are apparent in hypomethylation of intergenic regions (26% vs. 12% of hypermethylated probes). Our finding is consistent with a previous report in breast cancer, which detected the most marked hypomethylation in intergenic regions (Hon et al., 2012). Conversely, intergenic hypermethylation has been reported to be common in prostate cells, and such regions were shown to be evolutionarily conserved (Yegnasubramanian et al., 2011). In general, CpG methylation occurs in intragenic and intergenic non-coding regions far more commonly than in CpG island in 5'-promoters (about 3% of methylation occurs in CpG islands) and seems to regulate critical processes in a tissue-specific manner through alternative promoters (Maunakea et al., 2010).

It is important to be mindful of the array design, where 25% of probes correspond to intergenic regions, 41% to promoter regions, 31% to gene bodies and 3% to 3'-UTR regions, and many probes have multiple annotations. Thereby the most striking finding is that the majority of DMP correspond to gene body regions and intergenic sequences, rather than promoters, which supports the role of non-CpG methylation in the initial steps of the oncogenic process. Overrepresentation of intergenic and gene body probes may also be due to genetic aberrations, which are far more common in AK (as described in Chapter 3), or genomic instability.

6.4.4. Positional gene signature analysis reveals potential hotspots for epigenomic dysregulation in AK

Positional gene signature analysis revealed chromosomes 1, 6, 10, and 16 as containing potential “hotspots” for differential methylation. Additionally, hypermethylated genes were common on chromosomes 17 and 19, while hypomethylated genes occurred commonly on chromosomes 5, 7, and 20. Hypomethylation of cancerous genome has been associated with genomic instability and shown to preceed genomic damage (Rodriguez et al., 2006), thus these regions may represent areas of susceptibility to genomic alterations. Although a part of the positional signature for hypermethylated genes overlapped with regions of genomic gain in AK, copy-number variation has been previously shown to have minimal if any effect on methylation (Houseman et al., 2009). Given that 70% of samples were paired, germline mutations are also unlikely to underpin this finding.

6.4.5. Several oncogenic pathways are dysregulated in genes differentially methylated in AK

KEGG pathway analysis of hyper- and hypomethylated genes revealed that several oncogenic pathways are dysregulated among differentially methylated genes in AK. MAPK signaling pathway has been previously implicated in cSCC oncogenesis, in addition to Notch, TGF-beta, Wnt and mTOR pathway. Several pathways involved in cell-cell interaction were also dysregulated (Tight junctions, Focal adhesion, Gap junctions, Regulation of Actin cytoskeleton, EMC receptor interaction), all known to be dysregulated among genes driving EMT (Groger et al., 2012).

Hypomethylated genes showed over-representation of genes belonging to several metabolic pathways, which is concordant with expression data profiling. Hypermethylated genes, on the other hand, showed dysregulation of additional cancer-related pathways such as p53, Jak-STAT, PPAR and VEGF pathway. This indicates that methylation differences are not entirely stochastic in nature, yet can initiate distinct processes based on changes in methylation levels.

6.4.6. Poor correlation between gene expression and methylation is concordant with previous findings

Our data confirm that the correlation between gene expression levels and methylation is overall rather poor. Although the expression levels generally increase with decreasing methylation, the changes are not linear, with a wide spread and many outliers. As described above, hypermethylation of 3'UTR region increases gene expression, while hypermethylation of the body and the promoter leads to decreased gene expression. The association between promoter hypermethylation and transcription repression has been observed before (Herman and Baylin, 2003). Gene body hypermethylation, on the other hand, has been previously linked with gene transcription (Rakyan et al., 2008). This suggests that the relationship between methylation and gene expression is likely influenced by additional factors such as chromatin features or sequence variations (Wagner et al., 2014).

6.4.7. Transcription factor analysis shows no correlation between TF dysregulation or promoter status of genes targeted, but confirms significant association between gene promoter methylation independent of TF regulation

The purpose of this *in silico* analysis was to identify putative TF involved in AK onset. Our findings confirm that promoter hyper- or hypomethylation determines expression levels of a gene, that TF overexpression possibly influences target transcription (p value=0.05) and that significant over-representation of motifs for a transcription factor does influences gene expression levels of targets of FOXF2, PITX2, RUNX1 and SMAD3.

FOXF2, PITX2 and RUNX1 are up-regulated in AK, while SMAD3 is down-regulated in AK. SMAD3 is a member of the TGF-beta pathway, and plays a fundamental role in TGF-beta induced repression, namely repression of c-myc (Frederick et al., 2004). The expression of FOXF2 is regulated by Hedgehog signalling (Jeong et al., 2004) due to Gli-binding domains (Madison et al., 2009) and the gene is known to suppress WNT signalling during development and promote the production of extracellular matrix (Ormestad et al., 2006). Given that one of the ECM molecules with hypomethylated promoter and motif enrichment for FOXF2 overexpressed in AK is MMP13, and a critical Hedgehog pathway molecule PTCH1 is down-regulated in AK, I propose a direct mechanistic link between down-regulation of PTCH1, FOXF2 and MMP13 upregulation.

PITX2 is a homeobox gene that plays a critical role in vertebral left-right asymmetry (Campione et al., 1999). It was also shown to be overexpressed in thyroid cancer (Huang et al., 2010) and shown to activate canonical WNT signalling pathway in ovarian cancer cells (Basu and Roy, 2013). In our data, it may be responsible for loop-like upregulation of RUNX1 as this gene possesses hypomethylated promoter with motif enrichment for PITX2. This TF is known to play a fundamental role in hematopoiesis (Chen et al., 2009) and is involved in acute myeloid leukaemia onset (Miyoshi et al., 1991). In mouse skin, it was shown to promote hair follicle stem cells proliferation and the formation of chemically induced skin tumours (Hoi et al., 2010).

These results warrant further exploration of RUNX1 function in AK oncogenesis.

6.4.8. Histological subtype does not affect cSCC methylation profile

Our previous data show that methylation profile of cSCC cell lines is associated with tumour histological subtyping (see Chapter 5), which is concordant with results based on SCNA typing of cSCC (Purdie et al., 2009). Yet data obtained in this study did not find an association between cSCC histology and methylation profile. This may be explicable by the low number of poorly-differentiated samples, which seem to have the most distinct genomic and transcriptional profile. Alternatively, given that the samples were not microdissected and most likely contained additional tissue elements such as vessels, it may have diminished methylation differences that are driven by keratinocytes only.

One sample originating in a PUVA-treated patient showed distinct assoctiation of its methylation profile with “PUVA” treated as a phenotype in the association model. Long-term PUVA treatment is a known carcinogen due to induced DNA damage, yet the mechanism and character of DNA is different from that induced by ambient solar exposure (Nataraj et al., 1996) and may thereby be reflected in the methylation profile. While this finding may indeed be an array artefact due to having only 1 PUVA cSCC, it warrants further investigation of PUVA-induced skin tumours as an entity of independent aetiology.

6.4.9. Comparison of cSCC with normal skin and AK revealed more common hypermethylation of DMP

68% of DMP in cSCC compared to normal skin and 73% of DMP compared to AK were hypermethylated. While cancer cells are usually regarded as “globally hypomethylated” and hypermethylated in CpG islands compared to normal tissue (Esteller, 2007), recent study of an aggressive type of childhood leukaemia failed to detect global hypomethylation while observing expected promoter CpG hypermethylation (Basu and Roy, 2013). DMP corresponding to promoters were more commonly hypermethylated in cSCC (28% versus 24% compared to skin, 27% versus 22% compared to AK, p value<0.05), while global mean beta values were highly similar in all three tissues. Similarly, a study in renal cell carcinoma revealed widespread hypermethylation, predominantly in gene bodies (Hu et al., 2014) as observed in our data (discussed in greater detail in the following section). This observation may be due to the array design, or due to tissue and tumour specificity of methylation regulation.

6.4.10. Functional genomic distribution of differentially methylated probes in cSCC reveals more common hypomethylation in intergenic regions, while gene body and 3’UTR are more commonly hypermethylated in comparisons of cSCC to both skin and AK

Functional genomic distribution in cSCC compared to normal skin revealed massive hypomethylation of intergenic regions. This observation held true for comparison of cSCC to AK. Hypomethylation of such regions has been previously observed in breast cancer cells (Hon et al., 2012), and represents a similar proportion as DMP in AK (26%). As in the skin versus AK comparison, gene body probes were more commonly hypermethylated in cSCC and in greater proportion (59% vs. 49% in AK), in addition to promoter and 3’UTR regions. As discussed in the previous section, frequent gene body hypermethylation was recently described in renal cell carcinoma (Hu et al., 2014) and may represent a novel area for further exploration in cancer epigenetics in addition to promoter methylation. Given that both comparisons (skin vs. AK, skin vs. cSCC) were conducted with one set of normal skin samples, it may have introduced statistical bias. Additionally, the majority of samples in the skin vs. AK comparison were paired in terms of patient origin, while the majority of cSCC samples originated in a different patient cohort, thus potential genomic differences may be contributing to this apparent difference in functional genomic distribution of DMPs.

6.4.11. “Open Sea” probes are more commonly differentially methylated in AK DMPs compared with cSCC DMPs

An apparent difference between probes differentially methylated in AK and in cSCC compared to normal skin was in the percentage of probes that correspond to Open Sea and Island regions. Open Sea probes were far more common among DMP in AK (49% of hyper- and 57% of hypomethylated probes in AK, 38% and 35% in cSCC, respectively). Additionally, CpG island probes represented 14% and 11% of hyper- and hypomethylated probes in AK, in comparison with 29% and 32% in cSCC (p value<0.0001). This observation indicates that important differences in methylation occur predominantly in the CpG island and adjacent regions (Island and Shore constitute over 50% of both hyper- and hypomethylated probes in cSCC), while in AK, changes in methylation occur primarily in Open Sea regions (49% and 57% of hyper- and hypomethylated probes). Given that the comparison of cSCC and AK revealed minimal differences in methylation and DMP corresponded mainly to Open Sea regions, in addition to statistical bias by comparing both to the same set of normal skin samples, from a biological perspective this may indicate that Open Sea methylation changes in AK are largely stochastic and temporary in nature, and will not necessarily lead to a subsequent malignancy, while methylation changes that occur in CpG island are more stable in nature and thereby associated with cancer development.

6.4.12. Pathway analysis of differentially methylated genes in cSCC reveals strong overlap with pathway analysis in AK

In our data, 83% of pathways dysregulated in hypermethylated and 87% of pathways in hypomethylated genes in cSCC overlapped with those dysregulated in AK. Given that 34.5% of hyper- and 49.3% of hypomethylated probes overlapped between the two datasets, this finding may be driven by a statistical bias, but may also point to a molecular consensus in stage-wise development of cSCC, as it is likely that critical molecular changes that drive AK development are maintained during progression to cSCC.

6.4.13. Genes that maintain their differential methylation status in both AK and cSCC are regulated by polycomb group genes

In order to detect potential upstream regulators of genes that are differentially methylated in both AK and cSCC in the absence of our own ChIP-seq data, I compared these to curated datasets available in GSEA. This analysis identified MAPK8, SUZ12 and PRC2 as potential regulators of overlapping hypomethylated genes, and TP53, RB1, SOX2 and FOXP3 of overlapping hypermethylated genes.

MAPK8 (JNK1) has been shown to play a role in inducing EMT (van der Velden et al., 2012, Tiwari et al., 2013) but has also been described to function as a tumour suppressor in skin cancer (She et al., 2002). Given that MAPK8 is down differentially expressed in cSCC, overrepresentation of genes up-regulated upon its induction may point towards EMT induction in AK and cSCC onset. SUZ12 and PRC2 are polycomb group proteins, and increased levels of SUZ12 were described in cSCC cell lines compared to normal keratinocytes (Balasubramanian et al., 2010). PRC2 is important for the maintenance of hair follicle stem cells (Millar, 2011) and has been linked to lung cancer oncogenesis (Sato et al., 2013). Our data justify further exploration of the role of polycomb genes in AK and cSCC.

TP53 and RB1 are tumour suppressors and hypermethylation of their target genes may represent a mechanism of abrogating their tumour suppressor function. SOX2 has been recently described as a critical molecule in mouse cSCC oncogenesis (Boumahdi et al., 2014) and is significantly overexpressed in our cSCC data (p value 0.0007) but not in AK compared to normal skin (p value=0.4). Given that its putative target genes are hypermethylated in both AK and cSCC, it may be due to compensatory epigenomic changes to counteract its oncogenic effect. FOXP3 is significantly upregulated in AK (p value=0.02) but not in cSCC compared to AK (p value=0.9). This may indicate it is consistently upregulated early on in the oncogenic process. This is consistent with findings detecting increasing levels of FOXP3 in cutaneous and oral SCC in which it has been suggested to lead to immune escape of cancer cells (Schipmann et al., 2014).

6.4.14. Positional gene signature analysis detects overlap with regions of genomic imbalance in cSCC

Integration with regions of known genomic imbalance in cSCC detected a substantial overlap between genomic loci containing significantly more DMG in cSCC. This finding indicates that an important part of observed differences in methylation may be due to genomic changes. Conversely, this integration also identifies 12q13, 12q24, 15q22, 16q13, 16q22, 16q24, 17q21, 17q25 and 2p23 as potential hypermethylation, and 21q22, 2p25 and 6q25 as hypomethylation hotspots. While these areas may correspond to areas of genomic instability in this tumour set, it has been previously demonstrated that specific regions of the cancer genome are targeted for differential methylation in various tumours, and may represent tumour-specific epigenetic instability hotspots analogous to genomic imbalances (Lee et al., 2006a).

In our data set, genes with various cellular functions were detected in those hotspots: hypermethylation hotspot 12q13 contained tumour suppressor CDK2, keratin 71 and 80, in addition to CTDSP2, integrin-β7 and methyltransferase like 7a which were shown to be down-regulated in Burkitt lymphoma due to CD40 signalling (Hlavanda et al., 2007). 17q21 contains keratin 15 and 19 and histone deacetylase 5 (HDAC5) that has been demonstrated to play a critical role in the maintenance of pericentric heterochromatin (Fang et al., 2013).

Hypomethylation hotspot 6p25 contained c-Myc target 1 (MYCT1) which has been detected as an upregulated target of RUNX1-ETO fusion protein in acute myeloid leukaemia (Zhao et al., 2013a) and F-box protein 5 (EMI1) shown to be associated with poor prognosis in ovarian carcinoma (Budinska et al., 2013) if overexpressed, and overtranscribed due to HPV17 (Heinrichs et al., 2010).

6.4.15. Validation of 450K array with bisulfite-sequencing data

We used bisulfite-sequencing data (described in Chapter 2) to validate findings of methylation differences between skin and cSCC using Illumina 450K array. Although differences in methylation detected with bs-seq were less prominent namely in areas of less pronounced genomic differences, segregation in areas of higher methylation differences was sufficiently robust. Additionally, bs-seq data was obtained from pooled laser-capture microdissected samples, which may explain less profound differences in methylation. Moreover, previous comparison of Illumina 450K array and MeDIP-seq found a correlation of 0.68, a degree of overlap between hyper- and hypomethylated genes is thereby expected (Clark et al., 2012).

6.4.16. Comparison with 27K methylation data obtained in cultured cSCC cell lines detects only 21.3% of DMG

To examine cultured cSCC cell lines and normal human keratinocytes as valid models for exploration of methylation regulation in skin cancer, I compared 398 DMP detected in cultured cSCC with DMPs (p value 0.01) in cSCC clinical samples. While the methylation direction was consistent across all overlapping probes, the overlap was only little over one fifth of DMPs detected in cultured cSCC. This may be due to several factors: cultured cSCC included EB-derived cSCC which showed differences in epigenomic regulation compared to non-EB cSCCs (details are described in the previous chapter). Additionally, the passage number of some of the cultured cSCC cell lines was close to 20, and thereby the differences may be due to culture artifacts. Lastly, only three PHK were hybridised to the 27K array, and there may be a substantial statistical bias due to this low number of control samples.

6.4.17. Comparison of AK and cSCC reveals 104 differentially methylated genes

Perhaps the most unexpected finding was the absence of major methylation differences between AK and cSCC, given that our expression analysis revealed several hundred dysregulated genes. Furthermore, the overlap between DEG and DMG was only 20%. While this may be explained by different statistical approaches or the use of laser-capture microdissected samples for expression profiling, it is also important to stress the relatively low number of AK in the expression analysis (8 samples), whereas such power issues did not occur in methylation analysis. Additionally, using stringent quantile normalisation may have prevented detection of more subtle differences in methylation between AK and cSCC.

Interestingly, many of the hypermethylated genes play a direct role in regulation gene transcription. GSEA analysis also detected TP53 and and TP63 targets among hypermethylated genes, in addition to SMAD3 and SMAD2 targets.

Overall, these data indicate that the progression from AK to cSCC requires fewer changes at the epigenomic level compared to the progression from normal skin to AK, and that critical oncogenic processes are already initiated in AK.

6.5. Conclusions and future directions

This study shows that critical steps for the oncogenic process are initiated in premalignant skin lesions and lead to dysregulation of several pro-oncogenic pathways.

In addition to addressing the limitations of this study highlighted in the discussion, future steps include exploring differences in methylation based on increasing passage number to determine validity of this widely used and economical model in exploring cancer epigenomics. Additionally, more detailed analysis of bs-seq data will reveal genome-wide changes in methylation and identify areas of non-CpG methylation relevant in cSCC oncogenesis.

Additionally, detecting copy-number variation in skin, AK and cSCC samples from Illumina 450K intensity values (Feber et al., 2014), and comparing those to expression and methylation data would reveal which changes detected with the methylation array are due to genuine differences in methylation, and which are due to the gain or loss of genomic regions. Additionally, aberrant methylation in AK needs to be further validated with additional samples and techniques, such as pyrosequencing to confirm epigenetic dysregulation of genes from normal skin through AK to cSCC. Moreover, future statistical analysis should incorporate patient age at biopsy into the model to determine the effect of ageing in methylation in both normal skin and premalignant and malignant tissues.

To gain further insight into molecular regulation in cSCC, RNA-seq and chromatin immunoprecipitation sequencing would reveal whole transcriptome changes and relevant transcriptional targets in AK and cSCC onset, and integration with bisulfite sequencing data would identify gene regulatory networks in this malignancy and further characterise the interplay between the transcriptome and the epigenome in the progression from AK to cSCC.

7. MicroRNA profiling of normal skin and cutaneous squamous cell carcinoma

7.1. Introduction

As described in Chapter 1.5.5, microRNAs (miRNA) are small, highly conserved molecules that act as translational regulators at a post-transcriptional level (Ibanez-Ventoso et al., 2008). Given their role in regulating many fundamental cellular processes across species (Van Wynsberghe et al., 2011), it is perhaps not too surprising that a large body of evidence has accumulated demonstrating that miRNA dysregulation substantially contributes to the onset of majority of human malignancies (Croce, 2009).

The first miRNAs, lin-4 and let-7, were described in a nematode C. elegans (Lee et al., 1993, Reinhart et al., 2000) and were shown to regulate C. elegans development through negative post-transcriptional regulation by binding to 3′ untranslated region (UTR) of the cognate mRNA (Lee et al., 2002). Because miRNAs binding to 3’UTR is solely due to imperfect complementarity, miRNA are able to silence both tumour suppressors and oncogenes, thus playing an important role in the oncogenic process (Ventura and Jacks, 2009). Although the functional role of the majority of eukaryotic miRNAs has yet to be elucidated, genome-wide profiling of a variety of tumours has shown that miRNA profiling correlates with tumour type and differentiation (Lu et al., 2005), further stressing the relevance of this epigenetic mechanism in cancer.

MiRNAs are generated through a multi-step maturation process (delineated in Figure 7.1) that begins with the transcription of primary miRNA (pri-miRNA) by RNA polymerase II. This transcription product is several hundred to several thousand nucleotides long, and is subsequently digested into precursor-miRNA (pre-miRNA) by a microprocessor complex containing Drosha, a double-stranded RNA-specific endoribonuclease, and DGCR8, also known as Pasha, that functions as a molecular anchor capable of distinguishing the ssRNA-dsRNA junction (Han et al., 2006). Pre-miRNA is approximately 60-70 nucleotides long with a stem-loop structure (Lee et al., 2003b), and is transported from the nucleus into the cytoplasm by Exportin-5, a RanGTP-dependent protein. In the cytoplasm, Dicer (an endoribonuclease) forms a complex with TRBP, a ds-RNA binding protein, and this complex cleaves the terminal loop, leaving mature, double-stranded miRNA approximately 20 base-pair long. The so-called “passenger strand” is degraded, while the “guide strand” that is complementary to the target sequence through Watson-Crick base pairing is incorporated into the RNA induced silencing complex (RISC). This complex consists of Dicer, TRBP, protein kinase RNA activator (PACT), mature miRNA and Argonaut 2 protein (Redfern et al., 2013) that cleaves the complementary mRNA.

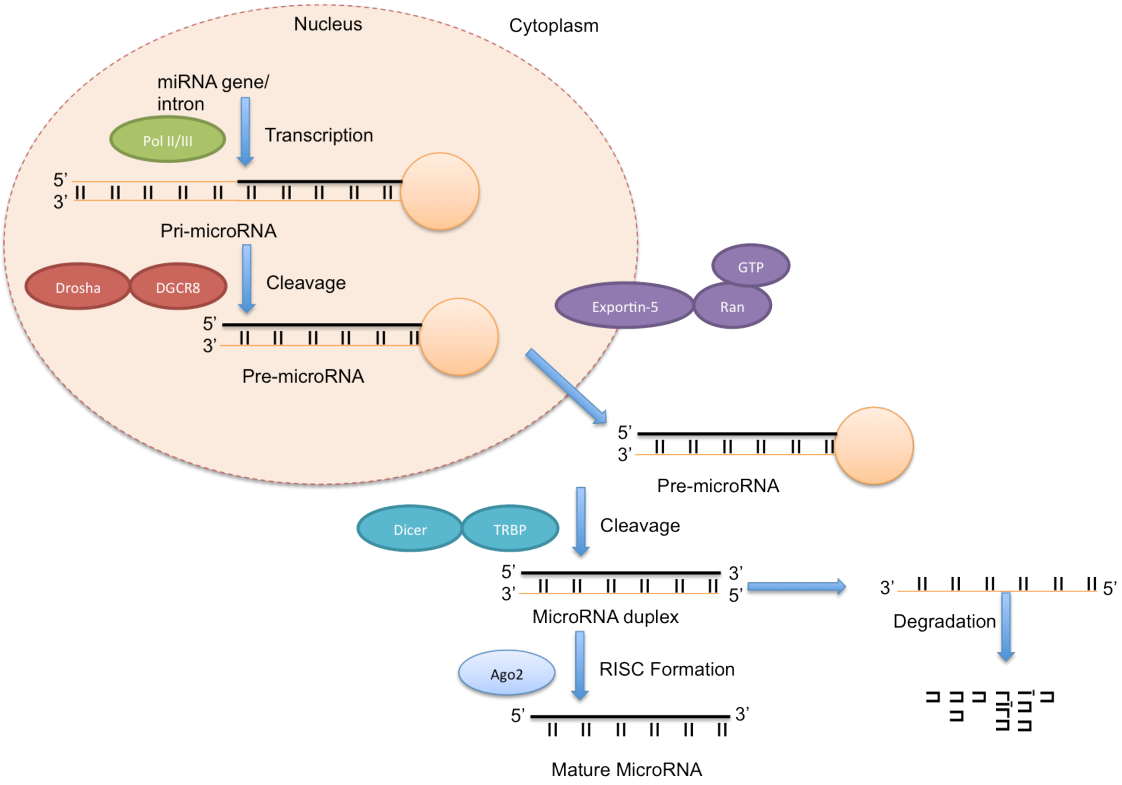


Figure 7.1 The process of miRNA maturation. Figure adapted from (Winter et al., 2009).

MiRNA profiling of the skin and diseased skin tissue to date has revealed an important role for miRNA in the development of cutaneous disorders. Several miRNA were shown to be dysregulated in dermatomyositis, including upregulation of miR-146b (Eisenberg et al., 2007), that has been implicated in various cancers including squamous cell carcinoma of the lung (Raponi et al., 2009) and papillary thyroid carcinoma (Yip et al., 2011). A study comparing miRNA levels in basal cell carcinoma and paired clinically normal tissue has found dysregulation of a total of 16 miRNAs (Sand et al., 2012b), in addition to miR-206 previously shown to play a role in BCC (Sonkoly et al., 2012). A summary of examples of miRNAs functionally involved in the pathogenesis of various skin conditions is provided in Table 7.1.

miRNAs	Putative target	Disease	Cell type	Up/ Down	Role	Reference
miR-29a	Type I collagen etc.	Scleroderma	Fibro-blasts	Down	Induces collagen synthesis and tissue fibrosis	(Maurer et al., 2010)
miR-196a	Type I collagen	Scleroderma	Fibro-blasts	Down	Induces collagen synthesis and tissue fibrosis	(Honda et al., 2012)
miR-223	PKCε	Dermato-myositis	Keratino-cytes	Down	Activates cell proliferation and leads to acanthosis seen in Gottron's papules	(Inoue et al., 2013)
miR-203	C-JUN	Basal Cell Carcinoma	Keratino-cytes	Down	Induces epidermal differentiation by stemness suppression	(Sonkoly et al., 2012)
miR-203	SOCS-3	Psoriasis	Keratino-cytes	Up	Activates proliferation and differentiation of	(Sonkoly et al., 2007)

					keratinocytes	
miR-424	MEK1 etc.	Psoriasis	Keratino-cytes	Down	Induces keratinocyte proliferation	(Ichihara et al., 2011)
miR-18a-5p	BCL2L10	Toxic epidermal necrolysis	Keratino-cytes	Up	Induces keratinocyte apoptosis	(Ichihara et al., 2014)
miR-146a	-	Melanoma	Tumour cells	-	MIRSNP G/C polymorphism regulates the proliferation, migration and invasion of tumor cells	(Yamashita et al., 2013)

Table 7.1 An overview summarising miRNAs involved in skin diseases (excluding cSCC). Table adapted from (Jinnin, 2014).

Previous studies exploring the role of miRNAs in cSCC have shown that miR-361-5p is decreased in cSCC and its decrease may play a role in VEGFA upregulation and subsequent tumour angiogenesis (Kanitz et al., 2012). MiR-124 and miR-214 that target ERK1 and ERK2, respectively, were shown to be downregulated in cSCC due to promoter hypermethylation (Yamane et al., 2013). A targeted approach has detected overexpression of miR-21 and miR-31 and down-regulation of miR-205 (Bruegger et al., 2013). Comparison of 4 cSCCs and 4 normal skin samples from healhy donors using TaqMan MicroRNA Low Density Array has revealed 58 differentially expressed miRNAs, and implicated miR-125b as a regulator of MMP 13 and potential tumour suppressor (Xu et al., 2012). Additional study of 7 cSCCs and 7 paired skin samples using Agilent miRNA microarrays found 13 upregulated and 18 downregulated miRNAs (Sand et al., 2012a).

The purpose of the current study was to detect differences in miRNA expression in a large series of paired cSCC samples and clinically normal skin from both immunocompromised and immunocompetent patients. Additionally, findings of differences in miRNA expression are paired with genome-wide expression and SNP data to detect transcriptional targets potentially regulated by miRNAs.

7.2. Materials and methods

7.2.1. Patients characteristics and tissue samples

Matched perilesional normal skin and tumour samples from a total of 27 patients were collected at the time of surgery and snap frozen at -70’C in liquid nitrogen. Additional normal skin from area distant to the tumour was collected from 3 patients. This set represented a new series of samples that were not previously used for genetic or transcriptional profiling.

Total RNA was isolated from a total of 52 samples, and 40 most appropriate samples (paired samples of normal and malignant tissue of the best RIN value) were were labeled with the miRCURY LNA™ microRNA Hi-Power Labeling Kit, Hy3TM/Hy5TM and hybridised to the miRCURY LNA™ microRNA Array (7'th GEN) (Exiqon, Denmark). Those 40 samples represented a total of 19 cSCCs, 20 perilesional skin samples and 1 distant normal skin, and were collected from 17 patients (5 females and 12 males), 13 of whom were immunocompromised (OTR) and 4 were immunocompetent (IC). Clinical details of samples are provided in Table 7.2.

Sample	**Histology**	**Gender**	**Immune status**	**Immune status**	**Concentration ng/µl**	**A260/A280**	**A260/A230**	**RIN**
VH Tumour	WD-cSCC	Male	RTR	RTR	92.80	2.06	1.78	6.2
VH Skin	Perilesional skin	Male	RTR	RTR	18.40	2.17	0.13	7.3
BD Tumour	MD-cSCC	Female	RTR	RTR	11.70	2.23	0.97	NA
BD Skin	Perilesional skin	Female	RTR	RTR	51.10	2.24	0.59	6.2
BR Lip Tumour	WD-cSCC	Male	RTR	RTR	49.60	1.98	1.29	3.5
BR Lip Skin	Perilesional skin	Male	RTR	RTR	37.60	2.11	1.75	2
BR Finger Tumour	WD-cSCC	Male	RTR	RTR	40.90	2.01	1.15	4.5
BR Finger Skin	Perilesional skin	Male	RTR	RTR	16.30	1.99	0.90	1
DT Tumour	cSCC	Female	IC	IC	64.60	2.10	0.30	6.1
DT Skin	Perilesional skin	Female	IC	IC	11.30	2.08	0.04	6.7
JD Tumour	WD-cSCC	Male	RTR	RTR	9.00	1.63	0.47	NA
JD Skin	Perilesional skin	Male	RTR	RTR	17.30	2.02	0.72	7.2
MF Tumour	MD-cSCC	Female	IC	IC	38.10	2.07	1.50	6.8
MF Skin	Perilesional	Female	IC	IC	27.40	1.43	0.58	NA

	skin							
TK Tumour	cSCC	Female	RTR	RTR	83.30	2.09	1.84	2.9
TK Skin	Perilesional skin	Female	RTR	RTR	8.90	1.81	0.56	5.7
VH2 Tumour	WD-cSCC	Male	RTR	RTR	132.60	2.07	1.87	NA
VH2 P Skin	Perilesional skin	Male	RTR	RTR	135.90	1.91	1.07	NA
VH2 Normal Skin	Distant normal skin	Male	RTR	RTR	12.90	1.68	0.12	2.4
WH Tumour	WD-cSCC	Male	IC	IC	185.10	2.07	1.99	4.2
WH Skin	Perilesional skin	Male	IC	IC	36.80	2.05	0.26	2.4
ML Tumour	MD-cSCC	Male	RTR	RTR	173.20	2.03	1.68	6.9
ML Skin	Perilesional skin	Male	RTR	RTR	38.00	2.11	1.36	6.9
CW Tumour	cSCC	Male	RTR	RTR	185.50	2.11	1.51	8.4
CW Skin	Perilesional skin	Male	RTR	RTR	96.40	2.14	2.12	NA
MB Tumour	MD-cSCC	Female	RTR	RTR	235.40	2.08	2.02	2.8
MB Skin	Perilesional skin	Female	RTR	RTR	149.80	2.07	1.84	2.6
JN Tumour	cSCC	Female	RTR	RTR	453.10	2.05	2.12	4
JN Skin	Perilesional skin	Female	RTR	RTR	81.90	2.02	1.22	6.5
PIN Tumour	WD-MD cSCC	Male	RTR	RTR	130.30	2.10	1.27	7.8
PIN Skin	Perilesional	Male	RTR	RTR	80.30	2.05	1.68	6.3

	skin							
AW Tumour	MD-cSCC	Male	IC	IC	698.80	2.10	2.19	2.3
AW Skin	Perilesional skin	Male	IC	IC	155.80	2.07	1.99	2.9
EC Tumour	MD-cSCC	Male	RTR	RTR	606.30	2.09	2.08	2.3
EC Skin	Perilesional skin	Male	RTR	RTR	156.30	2.09	2.00	2.6
PS Tumour	cSCC	Male	RTR	RTR	149.60	2.07	2.08	2.5
PS Peri	Perilesional skin	Male	RTR	RTR	57.00	2.06	1.96	5.6
PS NS	Distant normal skin	Male	RTR	RTR	69.50	1.69	0.83	4
BR NS	Distant normal skin	Male	RTR	RTR	118.00	2.06	1.94	2.7
JD2 Tumour	WD-cSCC	Male	RTR	RTR	140.60	2.08	1.63	2.9

Table 7.2 Samples hybridised to the miRNA microarrays and used for detection of differences between cSCC and clinically normal skin. If histology of cSCC samples was known, it was included in the histology section (WD=well differentiated, MD=moderately differentiated). RIN NA = RIN value below 1.

7.2.2. Exiqon microarrays

The miRCURY LNA™ microRNA Array uses locked nucleic acid (LNA)-modified oligonucleotides. LNAs are synthetic, thermostable RNA analogues with the ribose ring "locked" in 3-*endo* conformation by a methylene bridge between 2'-O and 4'-C atoms that allow miRNA profiling at uniform melting temperature. This unique feature allows the melting temperature for the array probes to be more uniform by varing the LNA content, and thereby allows for more specific miRNA hybridisation.

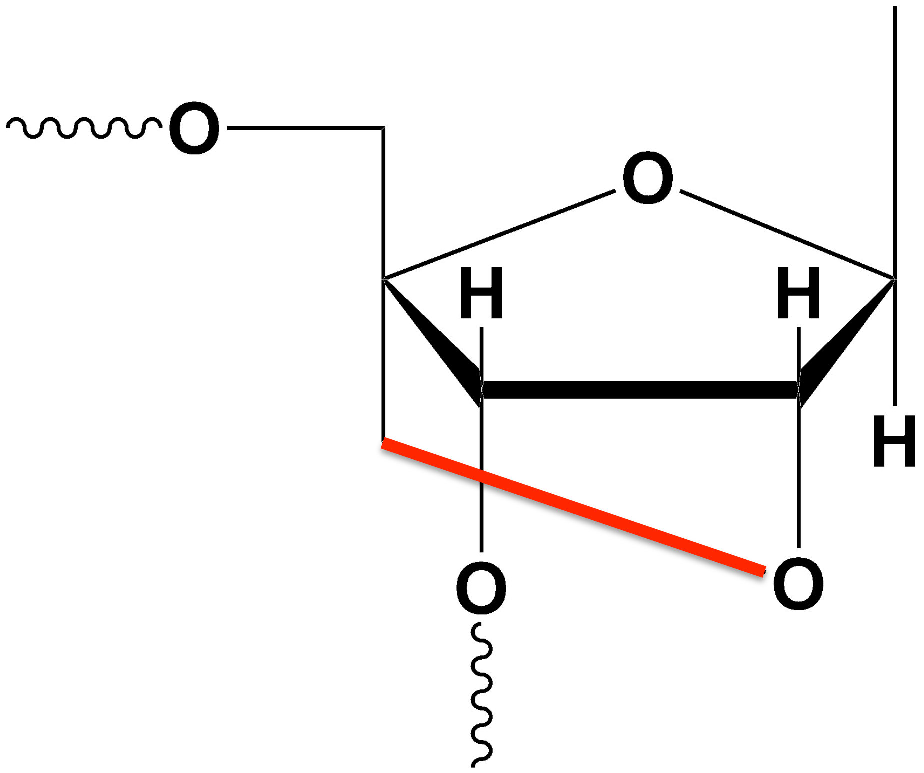


Figure 7.2 LNA nucleotide structure depicting the methylene bridge (in red) between the 4'-C and 2'-0 of the ribose ring. Figure adapted from http://www.vanderbilt.edu/vicb/DiscoveriesArchives/biomarking_infection_with_BSI.html.

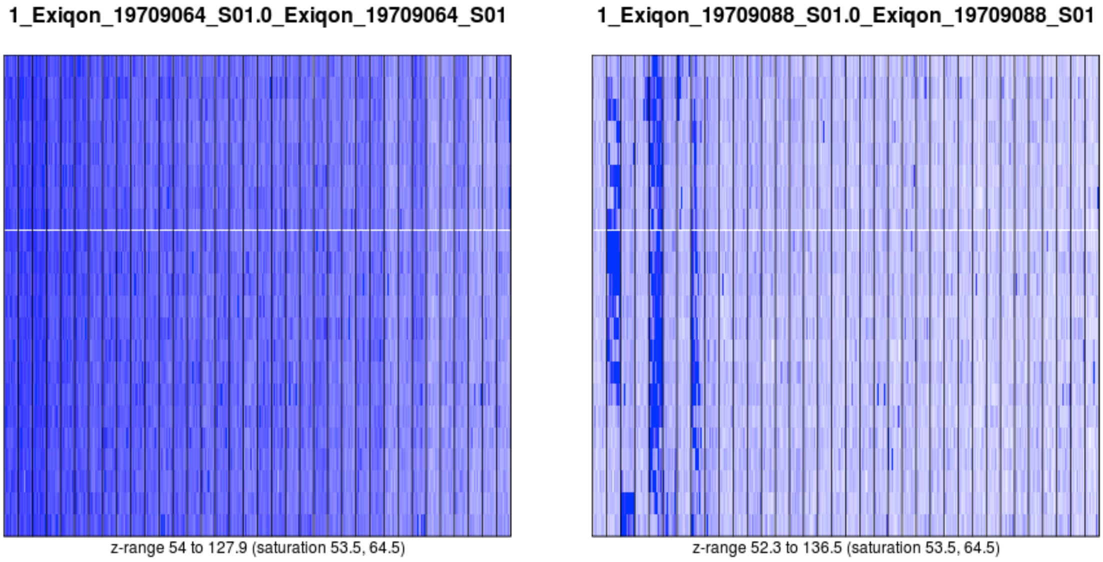
The array contains a total of 3100 capture probes that cover microRNAs annotated in miRBase version 19.0, in addition to relevant viral microRNAs and 25 proprietary miRNAs not listed in miRBase.

Array labelling and hybridisation was carried out by Exiqon (Denmark) according to the manufacturer's protocol. Following RNA quality control, samples were labelled by a two-step procedure with red and green dye and hybridised using Tecan HS Pro™ hybridisation station that automates both labelling and washing steps of the RNA processing. Briefly, one microgram of total RNA was mixed with spike-in miRNAs that are used for quality control and normalisation, and incubated with calf intestinal alkaline phosphatase (CIP) and CIP buffer for 30 minutes at 37°C in a PCR cycler with a heated lid. This process removes 5'-phosphates from RNA. Samples were then incubated at 95°C for 5 minutes and cooled on ice. Samples were then labelled with fluorophores by mixing 4 microliters of CIP reaction product with Hy3™ or Hy5™ fluorophores, labelling buffer, DMSO and a labelling enzyme and incubating the mix for 1 hour at 16°C in a PCR cycler, followed by incubation at 65°C for 15 minutes to stop the enzymatic reaction. Sample were then cooled at 4°C and hybridised to microarray slides within 2 hours.

Prior to hybridisation, equal amounts Hy3™ or Hy5™-labelled portions of a sample were mixed on ice and 25 µl of hybridisation buffer were added to the reaction followed by denaturation of the mixture at 95°C. Samples were then quickly cooled on ice, while microarray slides were prepared in the Tecan instrument by prewashing with buffer A for at 56°C for 30 seconds followed by flushing with 1x hybridisation buffer. 50 µl of previously prepared reaction mixture were then injected into the hybridisation station and incubated at 56°C for 16 hours while agitated at medium level. Following the incubation, samples were washed twice with Wash buffer A for 1 minute at 56°C, then for 1 minute at 23°C with Wash buffer C and finally for 30 seconds at at 23°C before drying.

All samples were processed in an environment with ozone control to prevent fluorophore degradation. Arrays were then scanned at 5 µm resolution with GenePix 4000A scanner (Axon Instruments, USA) and data were subsequently extracted and processed *in silico* by the author.

The raw data is captured as in image in ImaGene® files shown in Figure 7.3 and annotated in GAL files.



1_Exiqon_19709064_S01.0_Exiqon_19709064_S01

z-range 54 to 127.9 (saturation 53.5, 64.5)

1_Exiqon_19709088_S01.0_Exiqon_19709088_S01

z-range 52.3 to 136.5 (saturation 53.5, 64.5)

Figure 7.3 Example of ImaGene files. These files contain the raw data acquired from the microarrays for both the red and green channels.

7.2.3. miRNA microarray data analysis

Unless specifically stated otherwise, all figures and the entire analysis including quality control starting with the raw data was carried out by the author. The analysis was conducted using R, Bioconductor and ExiMiR package. Raw data stored in ImaGene TXT files were loaded into R and annotated using corresponding GAL file. One array failed data quality control and was excluded from further analysis (PS Tum, Figure 7.4). Raw values were normalised using RMA quantile normalisation without background correction.

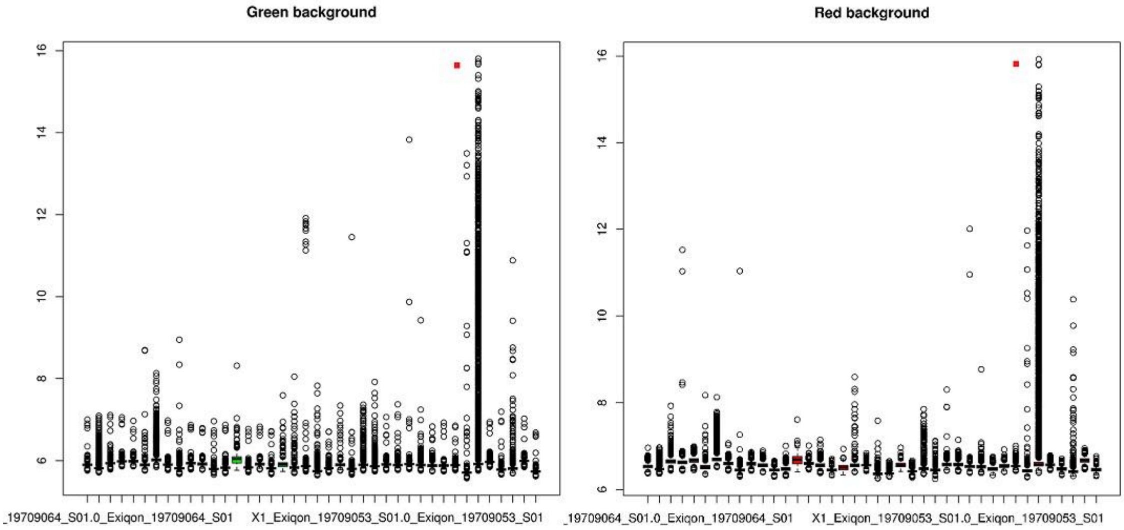


Figure 7.4 Boxplots of green (left) and red (right) background values of each individual sample (x axis) show a poor quality sample (labelled with red dot). Y axis=fold change.

MA plots show a correlation between log fold changes (M) and the mean single channel intensity (A). Gene plots are expected to overlap with spike plots in an ideal plot (Figure 7.5).

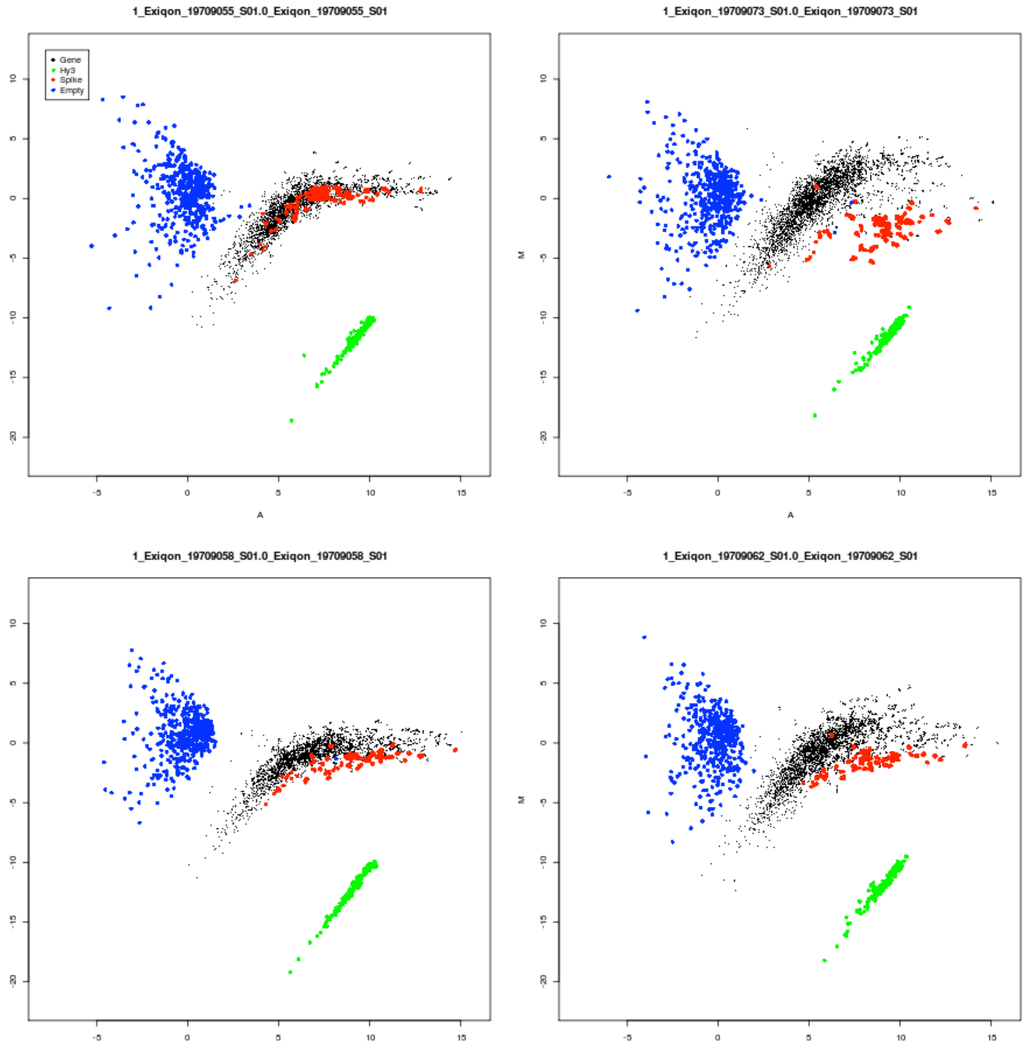


Figure 7.5 MA plot selected miRNA microarrays. Empty controls are shown in blue, spike-in controls are shown in red.

An offset equal to 10 was added to the normalised intensities before log-transforming in order to unify the lower intensities after log transformation towards zero (Figure 7.6). Offsetting is applies in order to stabilise the variance of log-ratios after transformation.

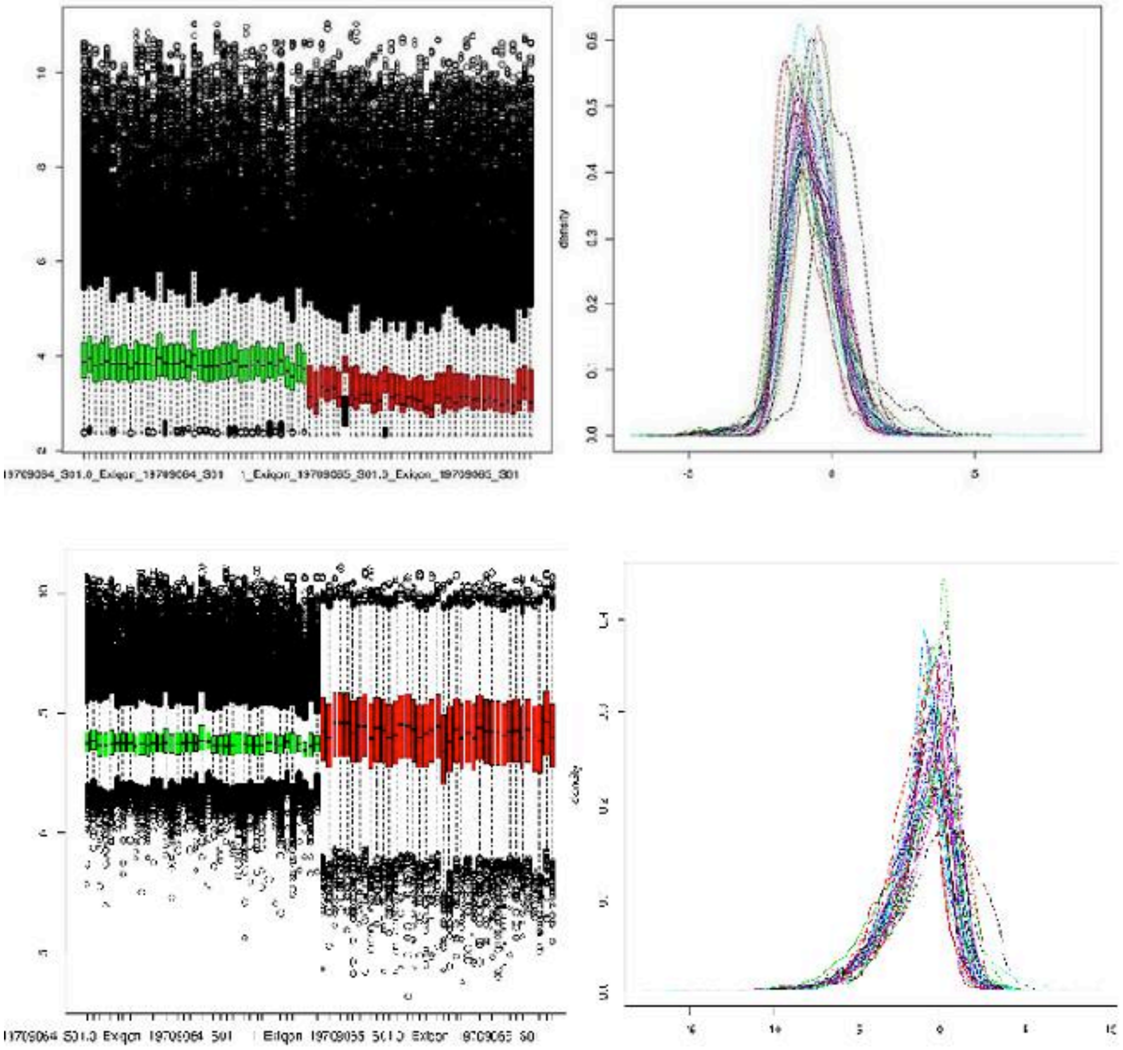


Figure 7.6 Normalised intensities before (lower panel) and after (upper panel) offsetting as demonstrated by boxplot and densityplot. As shown on the boxplot, "tail" of lower values is diminished or disappears in most samples after offsetting. This approach minimises the variation of intensities for very low intensities.

Correlation of samples and unsupervised hierarchical clustering was conducted on the normalised data using the VSN package in R using Pearson's correlation.

Differentially expressed miRNAs were detected using the lumi package (Du et al., 2008) with normalised, log2-transformed values. This package uses Bayesian modelling to detect differentially expressed elements and allows p-value adjustment for multiple testing (Benjamini-Hochberg correction).

Principal component analysis was conducted on all samples using log-transformed normalised values for 50 miRNAs with the highest standard deviation. The miRNA expression values were zero-shifted by subtracting the mean (the mean is zero), and the variance was scaled to equal 1.

7.2.4. Integration of copy-number variation loci in cSCC with miRNA data

Known genomic location of differentially expressed miRNAs was obtained in miRBase (http://www.mirbase.org). The genomic coordiantes of miRNA genes were then compared with coordinates of copy-number variation loci (CNV) described in Chapter 4. If the fold-change of the miRNA was concordant with the CNV (gain or loss), it was considered condordant with CNV.

7.2.5. miRNA target prediction and integration with genetic and transcriptional data

MiRNA target prediction was carried out *in silico* using Targetscan (Lewis et al., 2005). Although other algorithms were evaluated (DIANA, miRDB), it is not recommended to combine various target prediction approaches (Witkos et al., 2011) and only genes identified by Targetscan were used in subsequent analytical steps. This algorithm evaluates targets based on the presence of a conserved octamer or heptamer sites and their complementarity to the seed region of individual miRNAs. Additionally, targets that mismatch but compensate by conserved 3'-pairing are also included among predicted targets.

DIANA (DIANA)-microT-CDS as an online tool that predicts miRNA targets based on miRNA 3'-mRNA binding site complementarity and protein coding sequences (CDS) (Reczko et al., 2012). This algorithm uses existing high-throughput immunoprecipitation and sequencing data to generate a list of targets, and initial miRNA-mRNA hits based on 4-base pair-long complementarity in the seed region starting at seed position 1 or two, and then by the presence of G:U base pairs, single nucleotide bulges or mismatched sequences. This approach is less strict and inherently less precise than the approach utilised by Targetscan (Marin et al., 2013).

miRDB is an online annotated database using machine-learning approach that relies on Support vector machines (SVMs) to predict miRNA targets based on a large microarray training dataset (Wang and El Naqa, 2008). However, this program is limited by using only single 3'-UTR target sites for the training dataset rather than multiple target sites, since the presence of multiple targets for single miRNA can alter miRNA:mRNA interactions leading to low proficiency of gene repression (Garcia et al., 2011). This insight was incorporated in Targetscan prediction algorithm.

Predicted targets of all differentially expressed miRNAs which are expected to be down-regulated were then compared with differentially expressed genes (p value <0.01) between cSCC and AK (see Chapter 4 for details). Genes that showed expression changes consistent with genetic changes (described in Chapter 4) were excluded. The fold-changes of remaining genes (down-regulation or up-regulation) were compared with miRNA fold-change, and genes that did not show fold-change consistent with miRNA alteration (down-regulation if miRNA was up-regulated and vice-versa) were further excluded. The remaining genes were considered *bona fide* miRNA targets.

7.2.6. Validation of differentially expressed miRNAs

A subset of miRNAs that were detected as differentially expressed using the microarray were validated using three paired sets of cSCC and perilesional skin hybridised to the microarray and an additional set of 2 paired cSCC and adjacent normal skin samples listed in Table 7.3. Additional pairs are highlighted in italics.

Sample	**Histology**	**Concentration ng/µl**	**A260/A280**	**A260/A230**	**Gender**	**Immune status**	**RIN**
WH Tumour	WD-cSCC	185.10	2.07	1.99	Male	IC	4.2
WH Skin	Perilesional skin	36.80	2.05	0.26	Male	IC	2.4
MB Tumour	cSCC	235.40	2.08	2.02	Female	RTR	2.8
MB Skin	Perilesional skin	149.80	2.07	1.84	Female	RTR	2.6
JC T	cSCC	313.80	1.89	1.34	Male	RTR	8.1
JC S	Perilesional skin	44.50	2.00	1.84	Male	RTR	2.4
CW Tumour	cSCC	185.50	2.11	1.51	Male	RTR	8.4
CW Skin	Perilesional skin	96.40	2.14	2.12	Male	RTR	NA
BR Finger Tumour	WD-cSCC	40.90	2.01	1.15	Male	RTR	4.5
BR Finger Skin	Perilesional skin	16.30	1.99	0.90	Male	RTR	1

Table 7.3 A set of samples used for validation of differentially expressed miRNAs. If histology of cSCC samples was known, it was included in the "Histology" column (WD=well differentiated). RIN NA = RIN value below 1.

Validation was conducted using miRCURY LNA™ Universal RT microRNA PCR. Exiqon, Denmark carried out the total RNA processing and RT-qPCR. This approach utilizes pre-designed LNA™ PCR panel for detection of miRNAs with SYBR green, and uses spike-in control (UniSp6 CP) for normalisation. Only previously validated primers can be hybridised to the PCR plate, thus the selection of miRNAs for validation was limited by this technical aspect of the assay. The following 11 miRNAs were selected for validation: hsa-miR-142-5p, hsa-miR-1908, hsa-miR-210, hsa-miR-383, hsa-miR-423-3p, hsa-miR-498, hsa-miR-605, hsa-miR-628-3p, hsa-miR-638, hsa-miR-639, hsa-miR-943.

Total RNA was reverse-transcribed using the miRCURY LNA Universal RT miRNA PCR kit according to the manufacturer's incstructions in a single step. Briefly, 2 µl of total RNA (concentration of 5ng/µl) were mixed with synthetic RNA spike-ins, enzyme mix, 5x Reaction buffer and nuclease-free water to generate a total volume of 10 µl per reaction. Samples were then incubated at 42°C for 60 minutes, followed by 5 minutes incubation at 95°C to inactivate the reverse transcriptase and cooled on ice.

For the real-time PCR quantification, 20 µl of cDNA generated in the previous step was mixed with nunclease-free water and 10 µl of PCR mastermix containing SYBR green and added to qPCR 384-well plate pre-loaded with lyophilised primers with a pipetting robot. Each sample was tested in triplicate for each miRNA, and the plate contained also contained 4 blank controls containing no sample to detect contamination. The plate was then sealed and amplified with Roche Lightcycler 480® (Roche Applied Science, Germany). The reaction was carried out initially at 95°C for 10 minutes, then by 45 cycles amplified at 95°C for 10 seconds, followed by amplification at 60°C for 60 seconds with ramp-rate 1.6ºC/s cooling from 95°C to 60°C to ensure proper performance of the system.

7.2.7. RT-PCR data collection and quality control

The quality control consisted of melting curve analysis, amplification efficiency and a comparison of obtained Cp values to blank controls. Lightcycler 480® software was used to export raw Cp values and melting points for each sample. If several melting points were detected or if the melting point(s) were outside of the assay specifications (based on in-house database generated by Exiqon), these were removed from dataset and not included in the analysis. Additionally, reactions with amplification efficiency less than 1.6 and Cp values of 37 or more (blank controls have a Cp of 40) were also removed from the dataset.

7.2.8. RT-PCR data analysis

The raw Cp values (averaged for each sample) were normalised to averaged spike-in Cp values using the ΔCp method: average Cp of the target miRNA was subtracted from averaged Cp values of UniSp6 CP (ΔCp). This value was then transformed by exponentiating 2 to –ΔCp. This value was then log-transformed to detect fold-change.

Statistical analysis was performed using two-tailed, unpaired t-test using the GraphPad software.

7.3. Results

7.3.1. Correlation of global miRNA profiles detected in cSCC and adjacent normal skin samples

As shown in Figure 7.7, unsupervised hierarchical clustering of all 39 samples lead to a formation of several clusters largely based on their histology (cSCC and adjacent normal skin). While the separation of the samples based on histological typing is not complete, the fact that normal distant skin and skin adjacent to a tumour sample from the same patient clustered together (green and orange sample) shows that tissue miRNA profile is likely dependent on genetic background. In the right cluster, two tumour samples collected from the same patient (JN) cluster together, confirming this trend.

Additionally, the left-most cluster contains 13 samples of which 3 (2 skin samples, 1 tumour) are derived from a single patient (EC). The fourth sample from this patient (tumour) is within the rightmost cluster. Of the 10 remaining samples in the leftmost cluster, 2 pairs are from the same patient (BR, BD), and 2 more (TK and AW) have their paired tissue cluster right next to this set of 13 samples. Only one set of paired samples formed an adjacent cluster (CJ), which suggests that the majority of paired samples were sufficiently distinct in their miRNA profile.

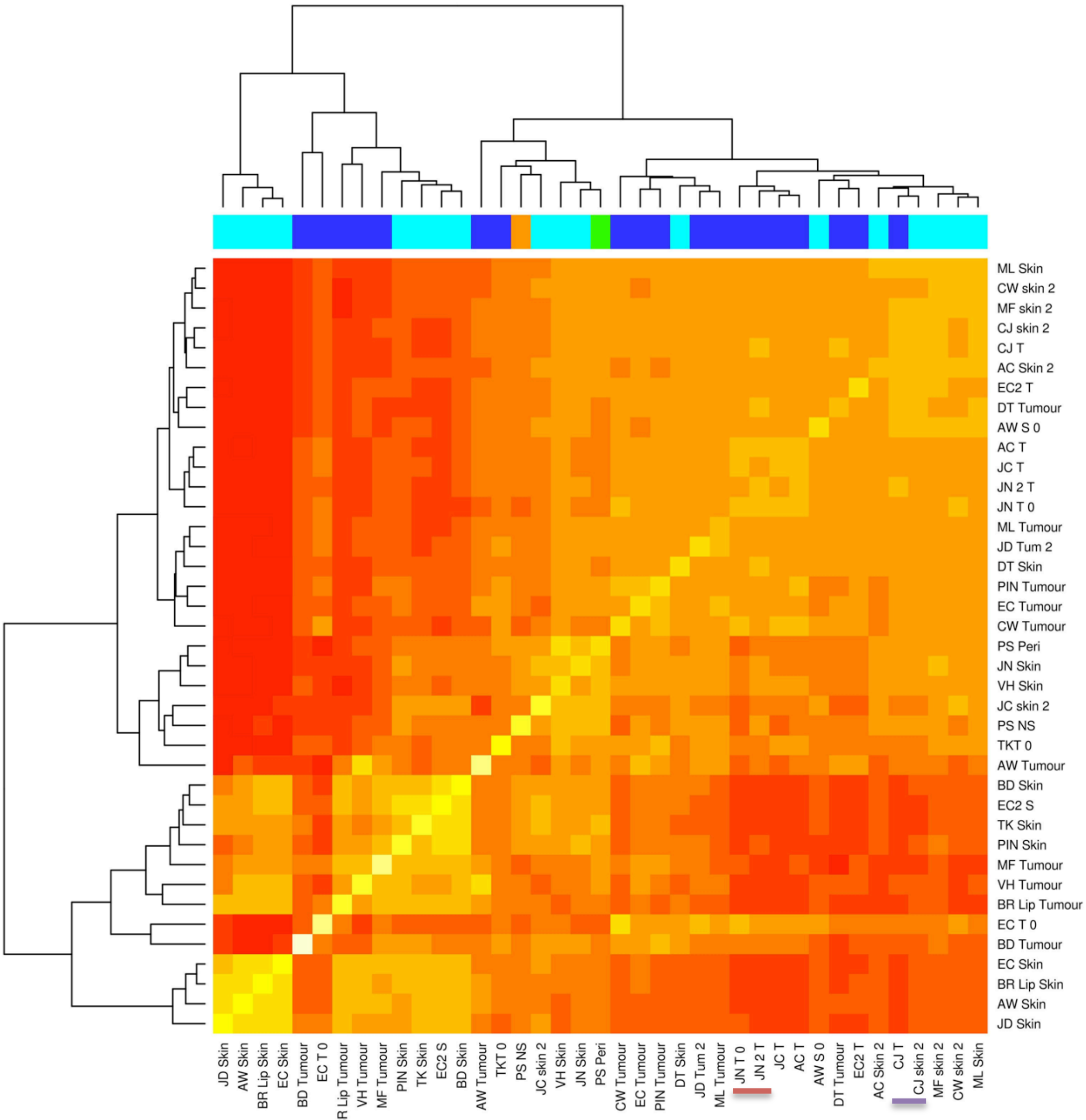


Figure 7.7 Unsupervised hierarchical clustering of miRNA profiles based on Pearson's correlation. Lightblue=normal adjacent skin, dark blue=cSCC, orange=distant normal skin, green=adjacent skin from a patient with distant normal skin. The red line indicates clustering of 2 different tumour samples obtained from the same patient. The purple line indicates clustering of paired tissue (tumour and adjacent skin) collected from one patient.

Principal component analysis (PCA) has shown that biological variation between the samples is sufficient to form biological classes in the majority of the samples (Figure 7.8). The plot indicated no biological differences between males and females, which was subsequently confirmed by comparison of miRNA tissue profile (i.e. cSCC and adjacent normal tissue combined) between males and females (data not shown).

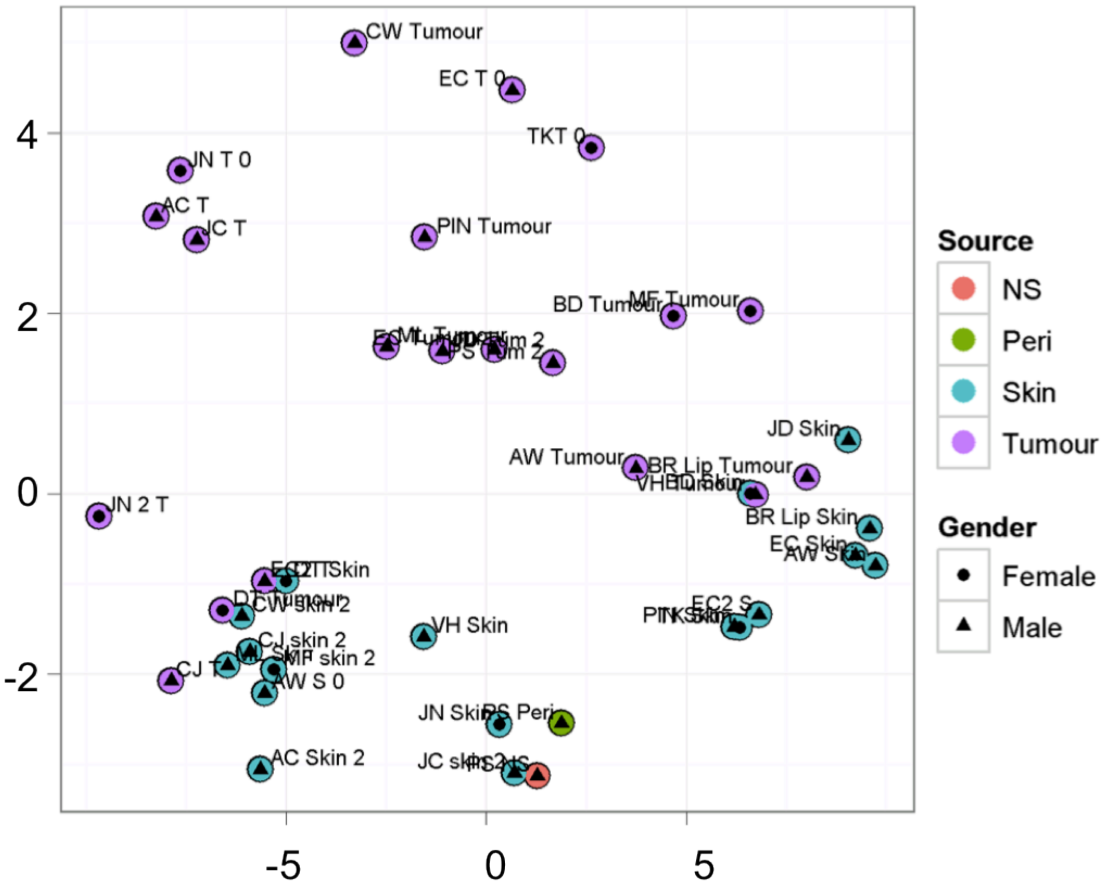


Figure 7.8 PCA analysis using 50 miRNA with the highest standard deviation. This plot shows good separation of the majority of tumours and clinically normal skin samples. X axis=PCA 1, y axis=PCA2. Figure prepared in collaboration with Exiqon, Denmark. Samples names correspond to samples listed in Table 7.2.

7.3.2. Comparison of miRNA profile between immunocompetent and immunosuppressed patients

Comparison of miRNA profile between all OTR and IC-derived tissues has not detected significantly differentially expressed miRNAs after p-value adjustment. Although a heatmap of top 15 differentially expressed miRNAs shown in Figure 7.9 indicates almost complete separation of OTR and IC-derived samples, it also indicates considerable variability of individual miRNA expression values across tissues, which explains the absence of significant findings.

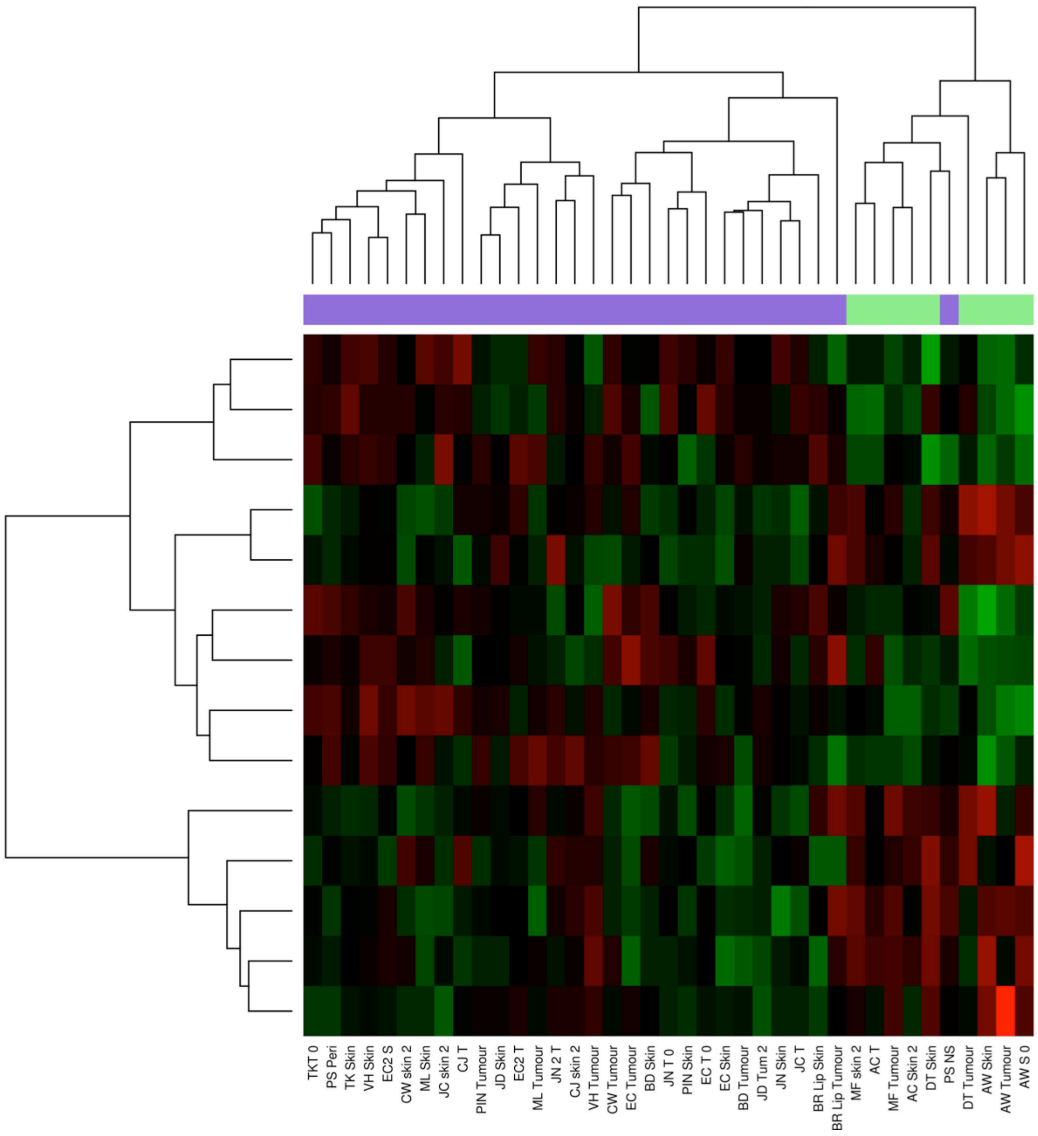


Figure 7.9 Heatmap of top 15 differentially expressed miRNAs (non-significant) shows separation of IC (green) and OTR samples. Green samples=IC patients, purple samples=OTR patients.

Comparison of both skin-only and tumour-only samples obtained from OTR and IC patients has not detected any significantly differentially expressed miRNAs after p-value adjustment. A heatmap of the top 17 differentially expressed miRNAs detected in comparison between tumours obtained from IC patients and OTR patients with p value of 0.01 or less before adjustment for multiple testing (Figure 7.10) shows similarly to the previous figure generally adequate separation of samples based on log2-transformed expression levels of those miRNAs, but the variability of expression across samples is apparent in this figure and underpins the absence of statistically significant findings.

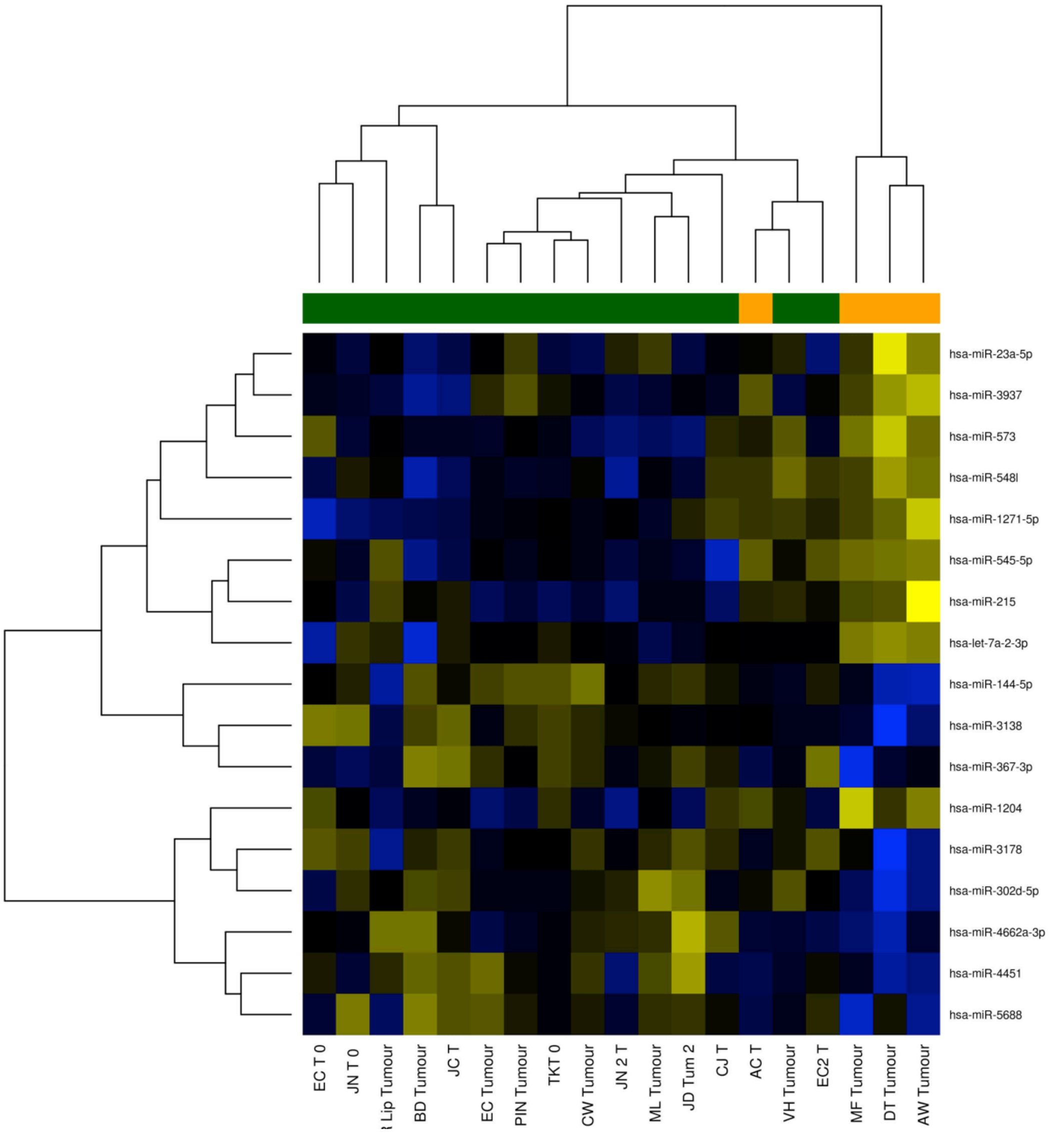


Figure 7.10 Heatmap of top 17 differentially expressed miRNAs (non-significant) comparing OTR-derived tumours with IC-derived tumours. Green=OTR tumours, orange=IC tumours.

7.3.3. Comparison of cSCC and adjacent normal skin miRNA profile

A comparison of all cSCC and skin samples has detected a total of 38 differentially expressed miRNAs using a stringent p value cut-off (p value <0.01), 6 of which showed fold-change of 1.5 or more, and 76 miRNAs using a less stringent cut-off (p value <0.05). Of those 76 miRNAs, 33 were upregulated and 43 were down-regulated (Figure 7.11). In the set of 38 miRNAs detected with more stringent p-value, 7 miRNAs were upregulated and 31 were down-regulated (Figure 7.13). The list of 38 differentially expressed miRNAs is provided in Table 7.4, and the full list of all 76 miRNAs is provided in Appendix 9.

Name	**logFC**	**P.Value**	**adj.P.Val**	**Chr**
hsa-miR-1469	-1.072505836	5.77436154519119e-10	1.20972874371755e-06	15
hsv1-miR-H17	-0.705890727	1.69129214680814e-09	1.77162852378153e-06	-
hsa-miR-548ap-5p/hsa-miR-548j	-1.137583927	6.37345196612178e-09	4.45079395634171e-06	-

hsa-miR-3656	-1.010574316	2.62424656124373e-08	1.3744491364514e-05	11
hsa-miR-4707-5p	-0.899529655	3.54495753467037e-08	1.48533720702689e-05	-
hsa-miR-371b-5p	-0.796315305	7.02535328882583e-08	2.45301919001502e-05	-
hsa-miR-638	-0.96944656	8.79507138186871e-08	2.63223922071642e-05	19
hsa-miR-4787-5p	-0.783080099	1.0748940076033e-07	2.81487868241115e-05	-
hsa-miR-4708-3p	-0.819929533	4.15585920532755e-07	9.6739167057347e-05	-
hsa-miR-4530	-0.912825789	5.83423681553376e-07	0.000122227	19
hsa-miR-4800-3p	-0.959652136	8.73797238933443e-07	0.000166419	-
hsa-miR-943	-0.547735401	1.53754766534614e-06	0.00026843	4
kshv-miR-K12-3-5p	-0.582550336	1.99976954076941e-06	0.00030663	-
hsa-miR-3960	-0.745469234	2.04908076517478e-06	0.00030663	9
hsa-miR-1909-3p	-0.625525465	2.71122173072946e-06	0.000378667	-
hsa-miR-4505	-0.965317513	7.52660351418713e-06	0.000985515	14
hsa-miR-744-5p	-0.691696164	9.56746704377922e-06	0.00117905	-
hsa-miR-642b-3p	1.141917999	1.15649194007355e-05	0.001275185	-
hsa-miR-605	-0.513822616	1.15575701021536e-05	0.001275185	10
hsa-miR-31-5p	1.740874987	1.25833940732504e-05	0.001318111	-
hsa-miR-21-3p	0.6724371	1.42998835672022e-05	0.001426584	-
hsa-miR-4732-5p	1.92248161	2.03475247609164e-05	0.001937639	-
hsa-miR-4423-5p	0.443625999	3.69901561629572e-05	0.003228932	-
kshv-miR-K12-6-3p	-0.793278205	3.57816180117349e-05	0.003228932	-
hsa-miR-585	-0.617744304	3.90871812017574e-05	0.003275506	5
hsa-miR-125b-1-3p	-0.349733408	4.82440516267739e-05	0.003609689	-
hsa-miR-3195	-0.694960429	4.49031145967249e-05	0.003609689	20
hsv1-miR-H7-3p	-0.708660707	4.66662449625971e-05	0.003609689	-
hsa-miR-495-5p	0.391768945	6.69281365222725e-05	0.004523047	-
hsv2-miR-H10	-0.713452548	6.68499148385216e-05	0.004523047	-
hsa-miR-663a	-0.749102069	6.44310090858787e-05	0.004523047	20
hsa-miR-3687	-0.865385123	8.8350480273067e-05	0.005784196	21
hsa-miR-4516	-0.697507752	0.000101523	0.006445202	16
hsa-miR-4734	-0.445995207	0.000112973	0.006961121	17
hsa-miR-124-5p	-0.298243171	0.000127269	0.007617956	-
hsa-miR-224-3p	-0.561808414	0.00013696	0.007970294	-
hsa-miR-1290	1.530172368	0.000147766	0.008366759	1
hsa-miR-639	-0.249764274	0.000178618	0.009847496	19

Table 7.4 List of 38 differentially expressed miRNAs in cSCC. Fold-change is relative to normal skin.

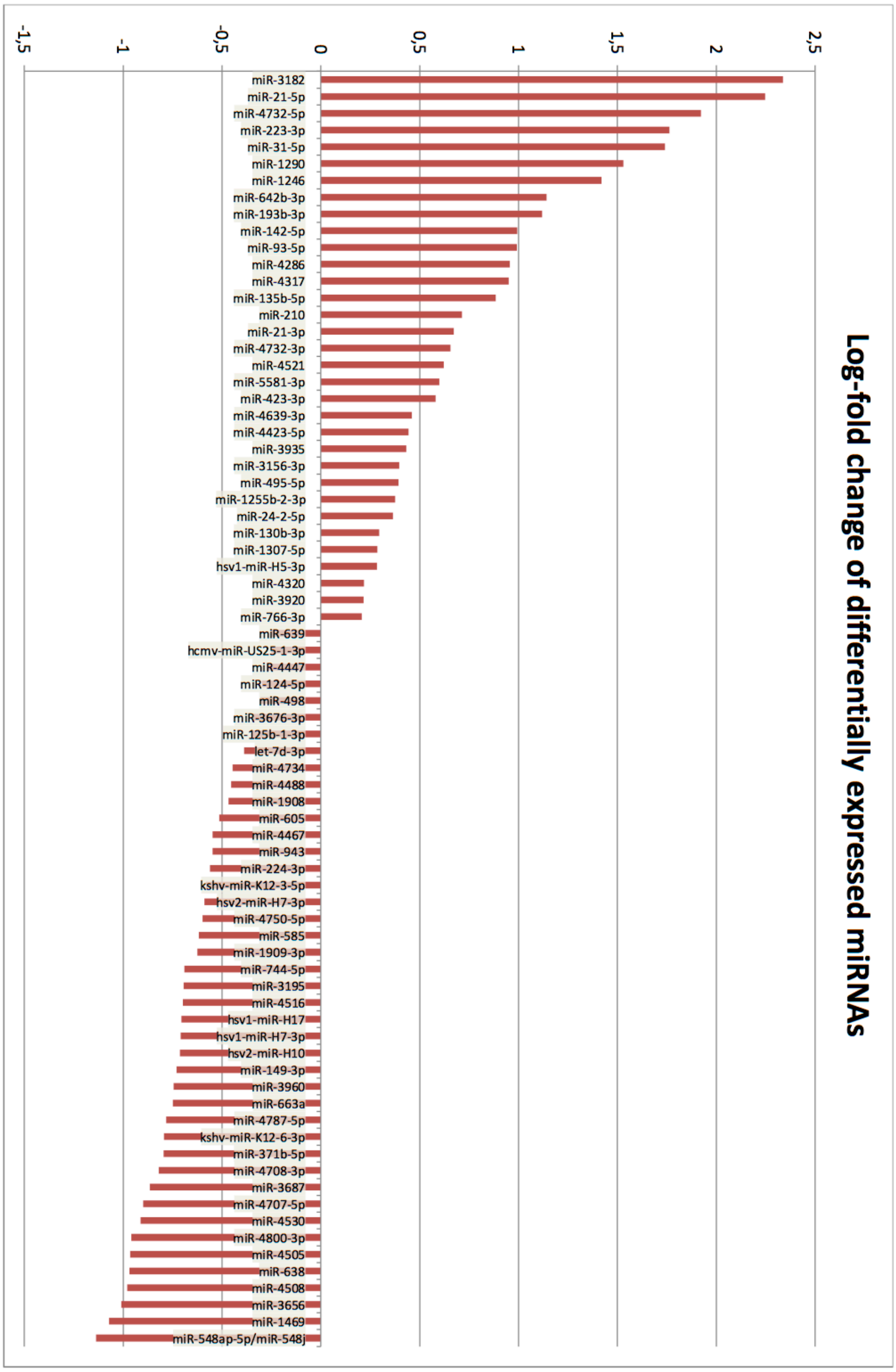


Figure 7.11 Bar plot showing fold-change of miRNAs detected with p-value <0.05.

Figure 7.12 Bar plot showing fold-change of 38 miRNAs detected to be differentially expressed in cSCC with p value <0.01.

A hHeatmap of 38 differentially expressed miRNAs is shown in Figure 7.13. With the exception of 3 skin samples, this set of miRNAs clearly separates skin from cSCC.

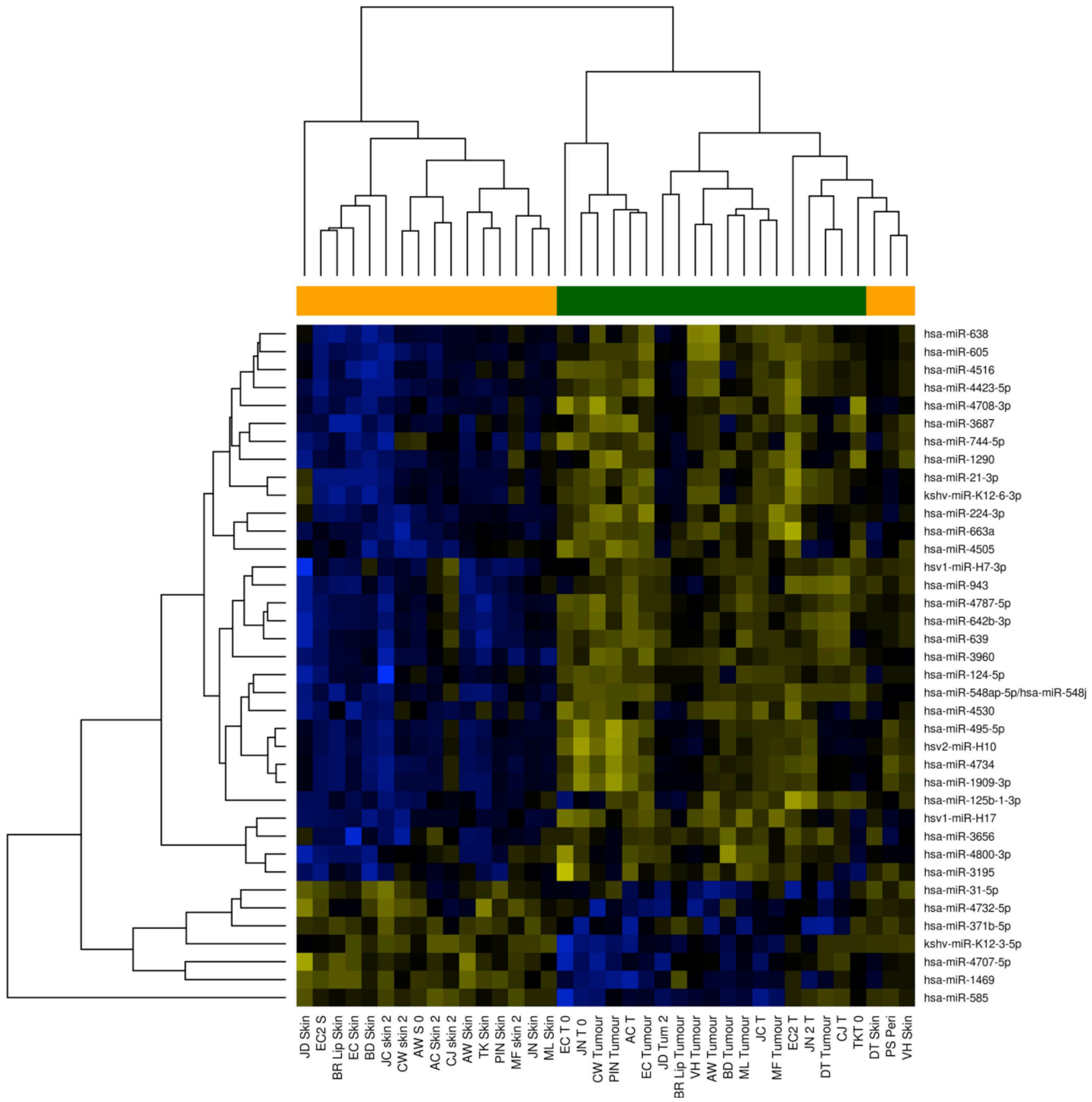


Figure 7.13 Heatmap of 38 miRNAs differentially expressed in cSCC. Orange=normal skin, green=cSCC.

7.3.4. Integration of miRNAs with genetic changes in cSCC

Previously published data that detected genomic changes in a large set of cSCC described in Chapter 3 and 4 were used for integrating known genomic location of 16 miRNAs (Figure 7.14).

Of those, 10 (62.5%) integrated with genomic changes (Table 7.5). Foldchange of 1 miRNA did not correlate with the expected genomic change (down-regulation in a region of genomic gain), and genes for the remaining 5 miRNAs were located outside of known copy-number variation loci in cSCC.

Chromosome	Location	Change	miRNA	FC
1	**1pter-1p32.3**	Gain	hsa-miR-1290	Up

4	**4pter-4p13**	Loss	hsa-miR-943	Down
5	**5qcen-5qter**	Loss	hsa-miR-585	Down
9	**9qcen-9qter**	Loss	hsa-miR-3960	Down
10	**10pter-10qter**	Loss	hsa-miR-605	Down
11	**11q14.1-11qter**	Loss	hsa-miR-3656	Down
16	**16p13.11-16pcen**	Loss	hsa-miR-4516	Down
17	**17pter-17qter**	Loss	hsa-miR-4734	Down
19	**19pter-19pcen**	Loss	hsa-miR-638	Down
19	**19pter-19pcen**	Loss	hsa-miR-639	Down

Table 7.5 MiRNAs that integrate with genomic changes in cSCC. FC=fold change.

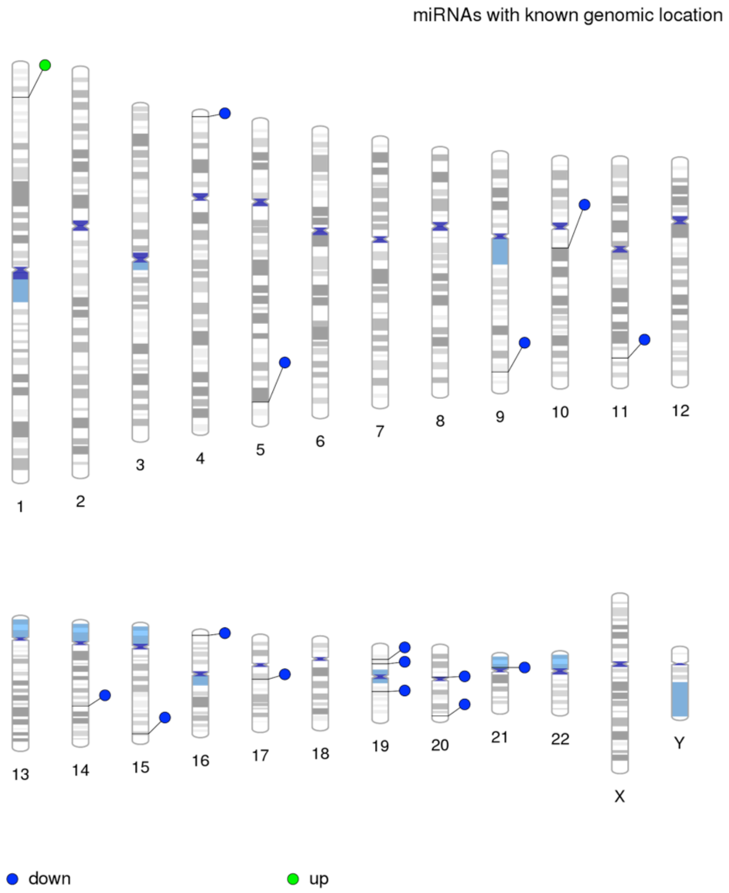


Figure 7.14 Genomic location of 16 miRNAs.

7.3.5. miRNA profiling of cSCC detects several viral miRNAs

The top set of 38 miRNAs contains 5 viral miRNAs, the full set of 76 miRNAs contained additional 3 viral miRNAs shown in Table 7.6. The majority of those were down-regulated with the exception of hsv1-miR-H5-3p. Five of those miRNAs were herpes simplex virus miRNAs, two were Kaposi's Sarcoma-Associated Herpesvirus miRNAs, and one human cytomegalovirus miRNAs.

Name	logFC	adj.P.Val	Comment
hsv1-miR-H17	-0.705890727	1.77162852378153e-06	Described in (Jurak et al., 2010)
kshv-miR-K12-3-5p	-0.582550336	0.00030663	
kshv-miR-K12-6-3p	-0.793278205	0.003228932	
hsv1-miR-H7-3p	-0.708660707	0.003609689	
hsv2-miR-H10	-0.713452548	0.004523047	Described in (Umbach et al., 2010)
hsv2-miR-H7-3p	-0.589493933	0.034912168	
hsv1-miR-H5-3p	0.283271184	0.040673136	
hcmv-miR-US25-1-3p	-0.251550024	0.048200154	

Table 7.6 Viral miRNAs detected to be differentially transcribed in cSCC.

7.3.6. Comparison with miRNAs previously described to be dysregulated in cSCC

Several miRNAs previously described as differentially transcribed in cSCC were detected in our dataset: up-regulated miRNAs included hsa-miR-21-3p and hsa-miR-21-5p (Bruegger et al., 2013), hsa-miR-210, hsa-miR-31-5p, hsa-miR-135b-5p (Gastaldi et al., 2014), hsa-miR-130b-3p and hsa-miR-766-3p, down-regulated miRNAs included hsa-miR-125b-1-3p (Sand et al., 2012a) and hsa-miR-124-5p (Yamane et al., 2013).

7.3.7. Detection of miRNA targets

Target prediction was carried out for the top set of 38 miRNAs. Various *in silico* tools utilising different algorithms for miRNA target prediction lead to a variety of targets, and the result of targets identified for 5 miRNAs by 3 different algorithms (DIANA, miRDB, Targetscan) is illustrated in Figure 7.15.

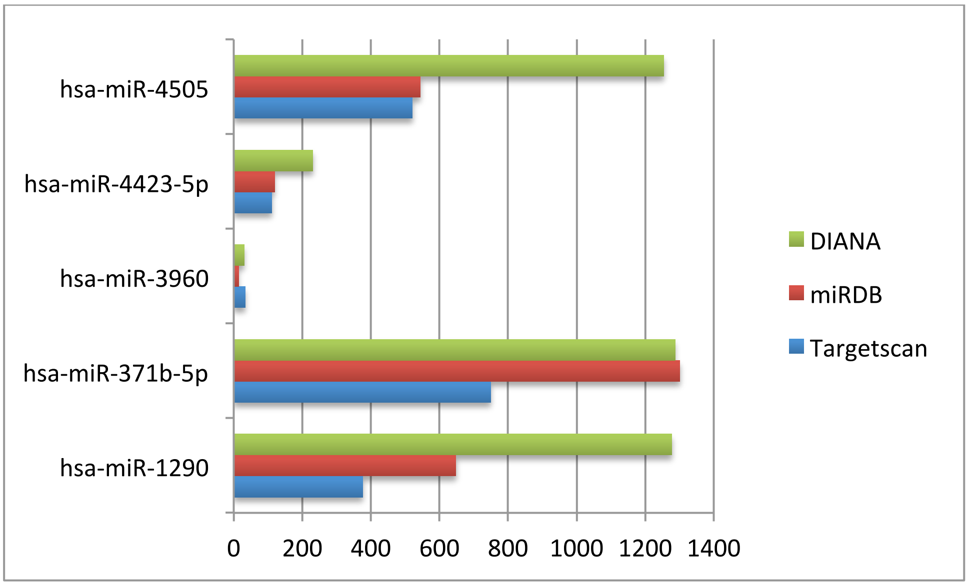


Figure 7.15 Number of genes predicted as targets of 5 miRNAs by three different algorithms.

A list of predicted genes and genes found in the list of differentially expressed probes generated with gene expression microarrays using fresh-frozen AK and cSCC samples is shown in Table 7.7. The AK and cSCC expression data are described in Chapter 3 and 4, respectively.

MiRNAs not listed have no predicted targets in Targetscan. Targets of only 12 miRNAs were identified to be differentially expressed in the expression dataset after subtracting genes that showed fold-change concordant with genetic changes and subsequently genes that did not show fold-change concordant with miRNA fold-change. This generated a final list of 18 miRNA targets listed in Table 7.8. Only one gene, TGFBR3, is targeted by 2 miRNAs.

Name	**Predicted**	**Found**	**%**	**Final list**	**% of found targets**
hsa-miR-371b-5p	749	15	2.002670227	1	6.7
hsa-miR-4530	547	17	3.10786106	5	35.3
hsa-miR-4505	521	7	1.343570058	2	28.6
hsa-miR-1290	376	12	3.191489362	3	25
hsa-miR-4708-3p	305	14	4.590163934	4	28.6
hsa-miR-4516	260	5	1.923076923	1	20
hsa-miR-605	260	11	4.230769231	2	18.2
hsa-miR-3960	257	1	2.941176471	0	0
hsa-miR-4423-5p	111	2	1.801801802	1	50
hsa-miR-943	101	1	0.99009901	0	0
hsa-miR-4732-5p	42	1	2.380952381	0	0
hsa-miR-638	30	2	6.666666667	0	0
hsa-miR-3656	21	0	0	0	0
hsa-miR-4707-5p	20	0	0	0	0
hsa-miR-4734	18	0	0	0	0
hsa-miR-639	15	0	0	0	0
hsa-miR-4787-5p	10	0	0	0	0
hsa-miR-3687	8	0	0	0	0

hsa-miR-3195	5	0	0	0	0
hsa-miR-585	5	0	0	0	0
hsa-miR-1469	2	0	0	0	0
hsa-miR-4800-3p	1	0	0	0	0

Table 7.7 List of miRNA targets predicted by Targetscan and the number of corresponding genes found to be dysregulated at the transcriptional level comparing AK and cSCC.

Name	**Number of targets**	**Target genes**
hsa-miR-1290	3	HS3ST1 TGFBR3 PDZD2
hsa-miR-4530	5	BMPR1B NRP2 CARHSP1 CYP27B1 UBE2H
hsa-miR-4505	2	EHD2 KLF6
hsa-miR-4708-3p	4	HN1 KLF7 VKORC1 N4BP1
hsa-miR-605	2	MAPKAP1 ETS2
hsa-miR-4516	1	ACTN1
hsa-miR-371b-5p	1	DENR
hsa-miR-4423-5p	1	TGFBR3

Table 7.8 Final list of miRNA targets.

Comparison of miRNA targets with differentially methylated genes identified in the previous chapter found no hypermethylated targets, and 2 genes (PDZD2, NRP2) were hypomethylated.

I hypothesised that the number of genes found in the expression list correlated with the number of predicted genes. This was confirmed by Pearson’s correlation (r=0.8956, p value<0.0001, Figure 7.16). Similarly, the number of predicted genes significantly correlated with the number of final targets (r=0.7539, p value <0.0001, Figure 7.17).

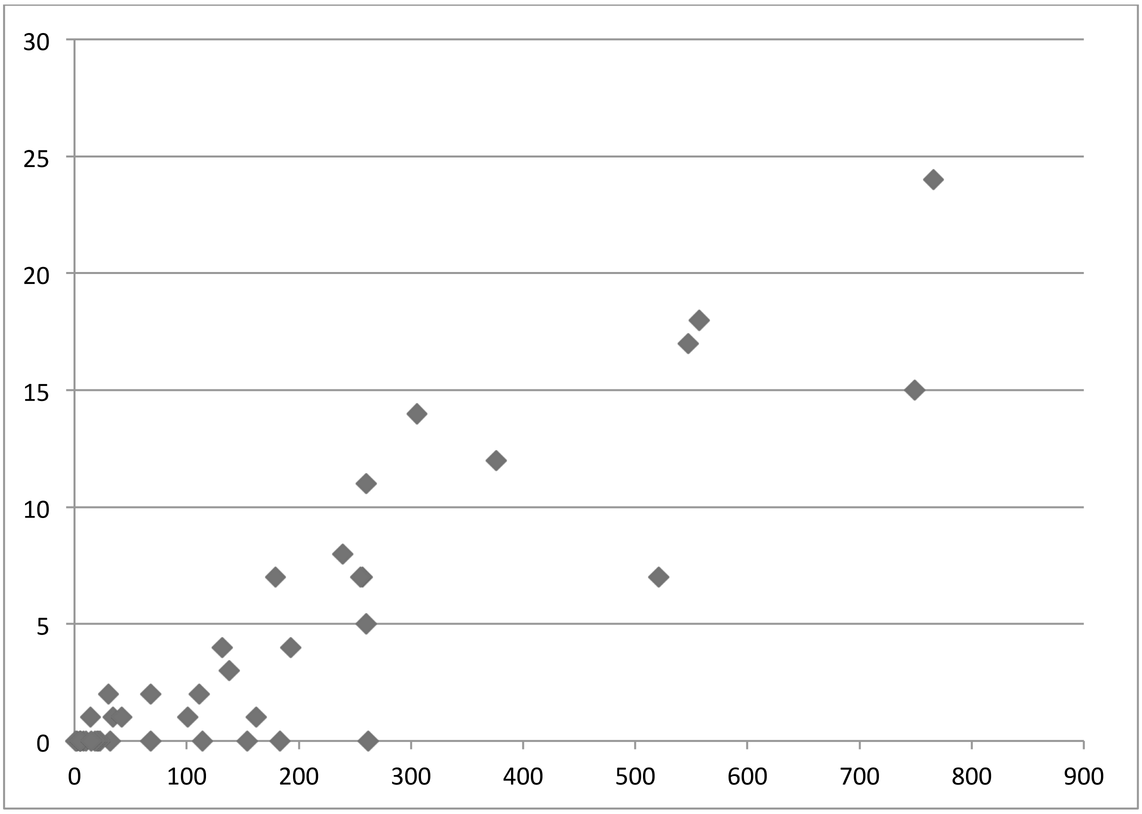


Figure 7.16 Scatterplot showing correlation between the numbers of predicted and detected miRNA targets. X axis=predicted genes, Y axis=genes detected in list of differentially expressed genes.

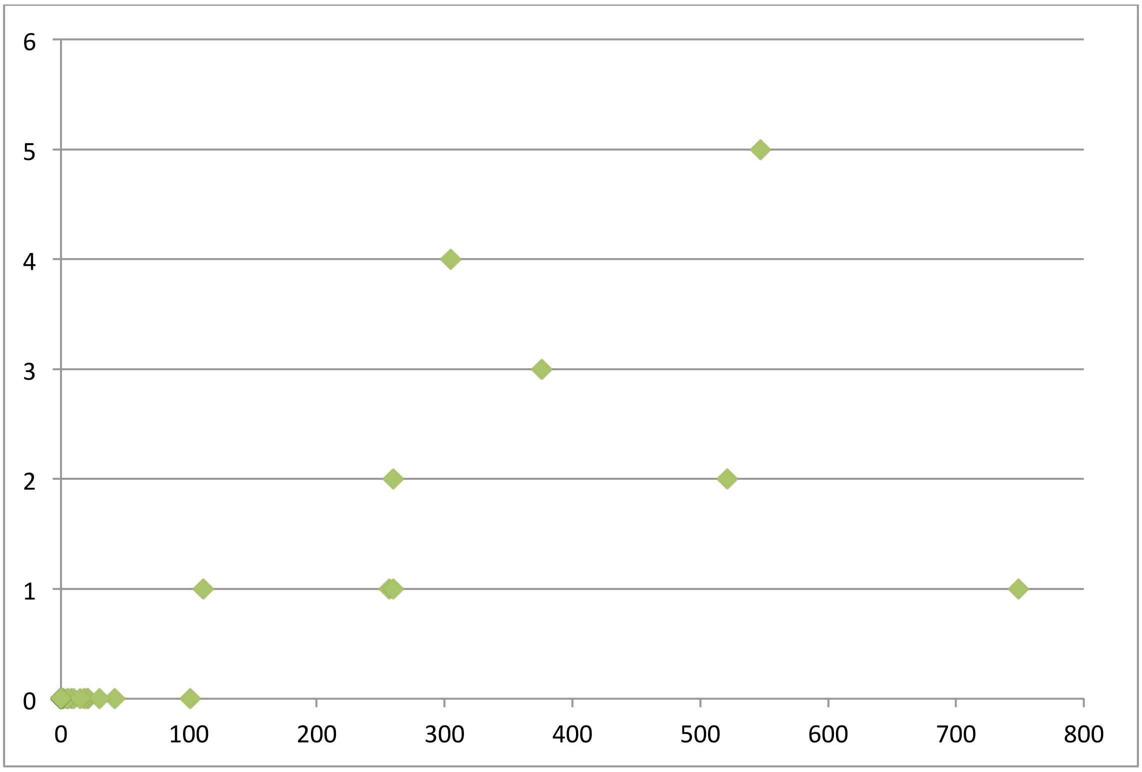


Figure 7.17 Scatterplot depicting correlation between the number of predicted genes and final miRNA targets. X axis=predicted genes, Y axis=genes detected in list of differentially expressed genes.

To compensate for the fact that miRNAs were evaluated in normal skin and cSCC, while the expression list used for target detection was generated by comparison of AK and cSCC, rather than normal skin, I used expression data generated with RNAlater-treated skin and AK samples (described in Chapter 3) and used average log2-transformed expression levels to observe general trends across sun-exposed skin (SE), non-sunexposed skin (NSE), actinic keratosis (AK) and cSCC.

As shown in Figure 7.18, with the exception of KLF6, the expression trend (over or underexpression in comparison with cSCC) in skin samples was concordant with changes detected in AK.

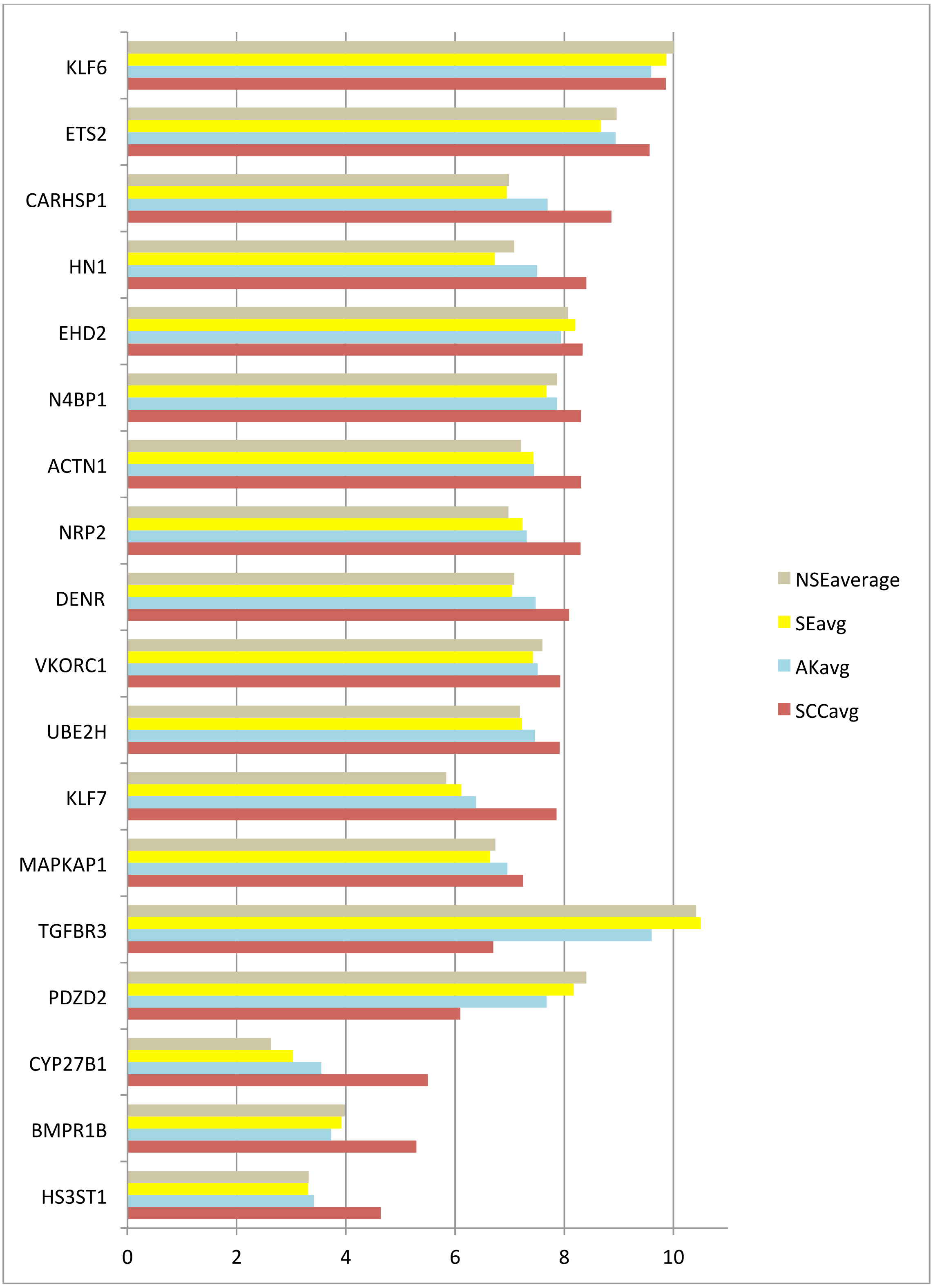


Figure 7.18 Average log2-transformed fold expression of 18 target genes as detected with expression arrays described in previous chapters.

To confirm this trend, I generated a heatmap of targets for hsa-miR-4530 and hsa-miR-4708-3p that were detected to have four or more targets. Expression levels across all five sample categories (cSCC, metastatic lymph nodes, AK, NSE, SE) of those targets lead to clustering trends concordant with histological typing.

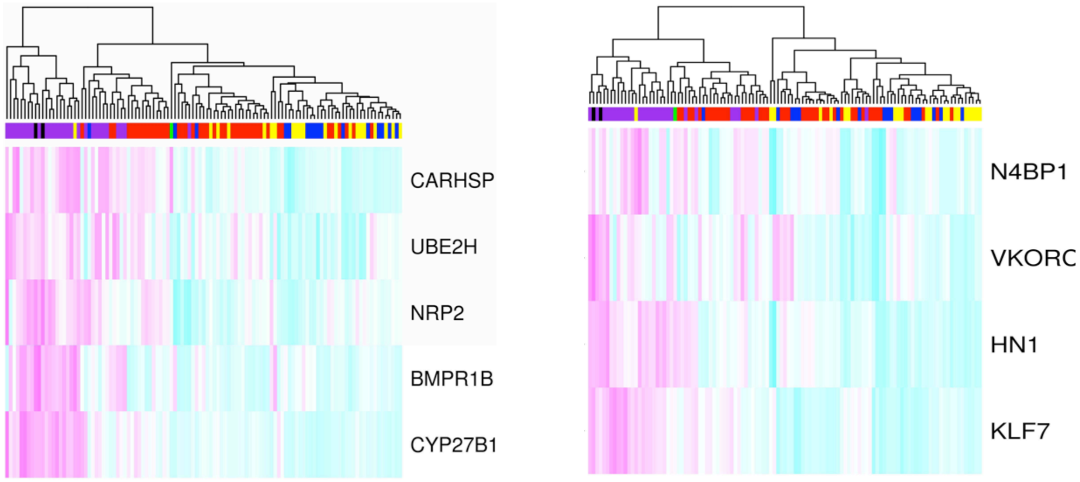


Figure 7.19 Heatmap of log2-transformed expression values for hsa-miR-4530 (left panel) and hsa-miR-4708-3p (right panel) targets. Unsupervised clustering of all samples lead to mostly uniform clusters of cSCC (purple), lymph node metastases (black), AK (red), NSE (blue) and SE (yellow).

Genomic location of all 18 targets along with expression levels in cSCC, AK, SE and NSE is shown in Figure 7.20. Three target genes are located on chromosome 16, two on chromosome 1, 4 and 12, and the remaining 9 targets were each found on 1 chromosome.

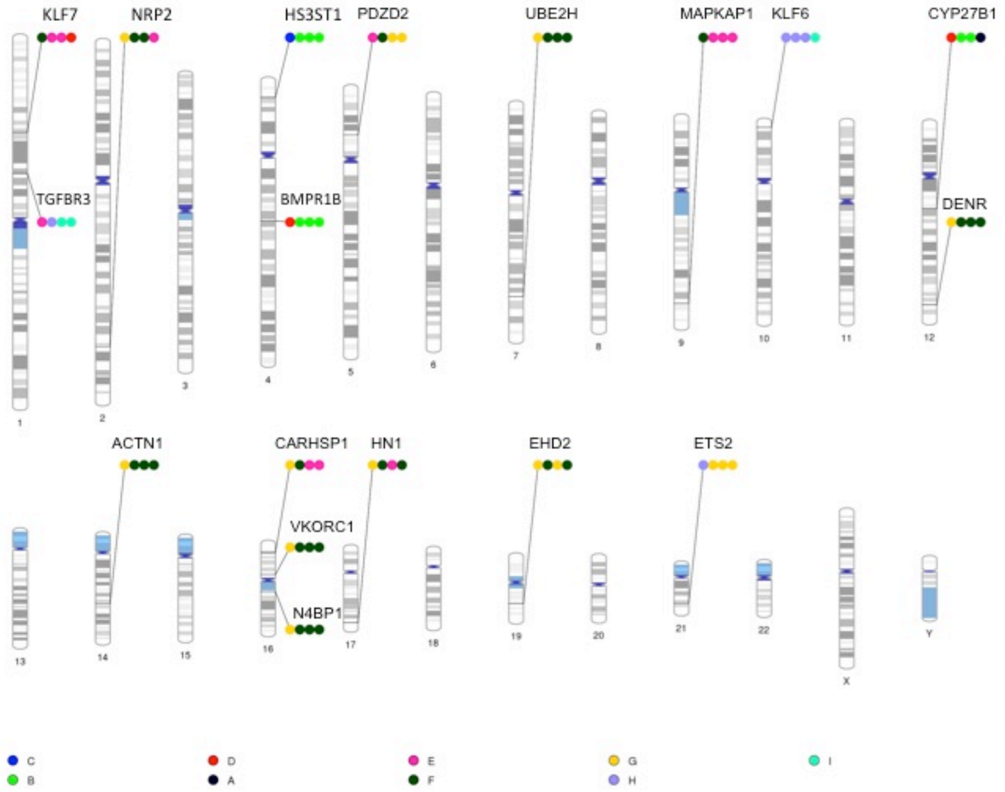


Figure 7.20 Genomic location of 18 miRNA targets. Innermost circle=cSCC, followed by AK, SE and NSE (outermost circle). Letters correspond to ranges of log2-transformed expression levels. A=2-3, B=3-4, C=4-5, D=5-6, E=6-7, F=7-8, G=8-9, H=9-10, I=10-11.

7.3.8. Validation of miRNA microarray

A novel approach to miRNA validation that uses lyophilised primers bound to wells of PCR plates was utilised to validate selected miRNA targets. Only 3 of 11 miRNAs were detected in all samples. Moreover, both the number and detected targets were often not equal across biological replicates (triplicates of the same sample).

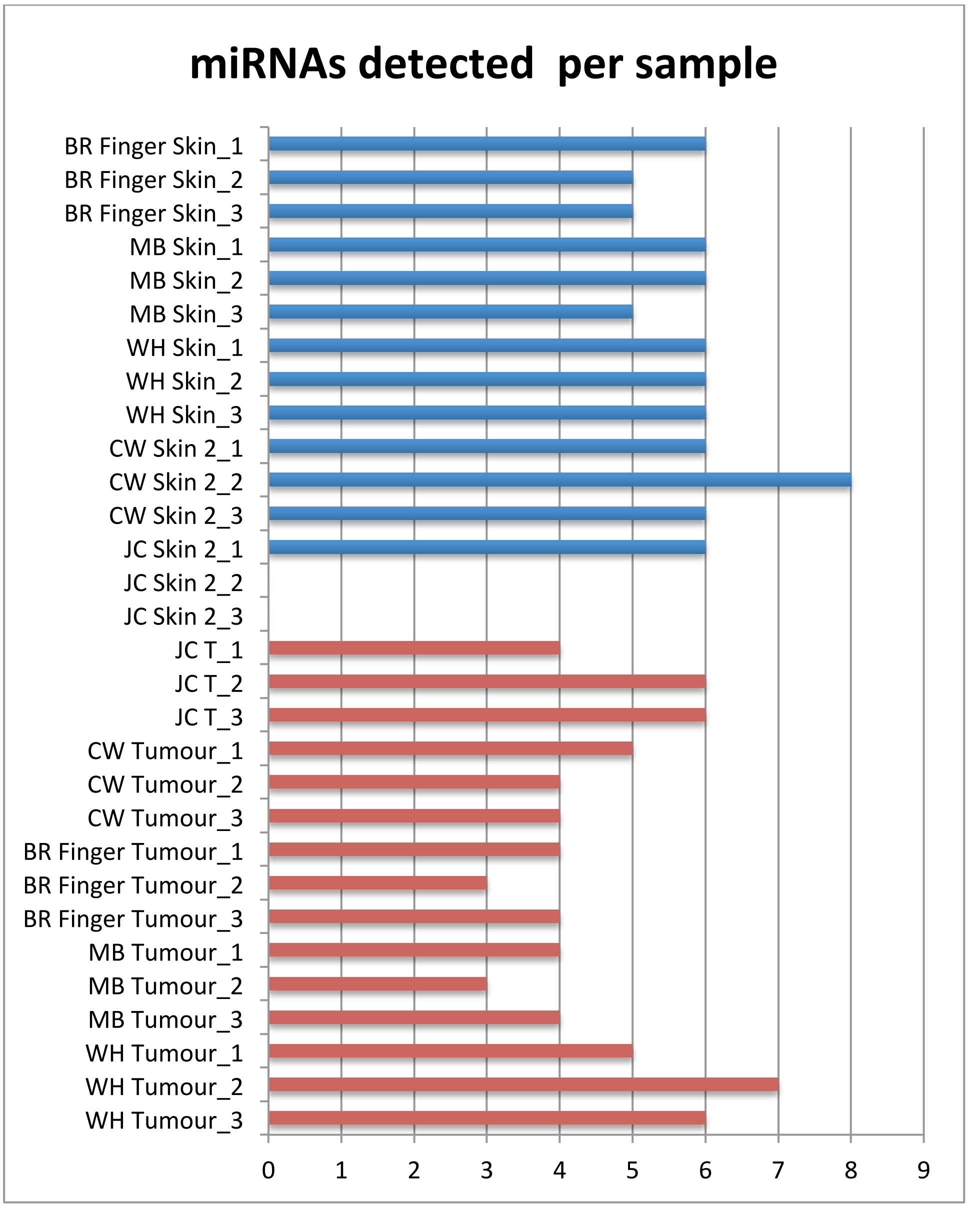


Figure 7.21 Number of miRNAs detected in each cSCC and adjacent normal skin sample in microarray validation. This figure shows lower miRNA detection in tumours samples.

Addtionally, 7 of 11 miRNAs failed quality control (hsa-miR-383, hsa-miR-423-3p, hsa-miR-498, hsa-miR-605, hsa-miR-628-3p, hsa-miR-638, hsa-miR-639) and were not analysed further. Of the remaining four miRNAs, 2 were significantly differentially expressed (hsa-miR-1908, p value=0.03, and hsa-miR-943, p value=0.0016) and concordant with fold-change detected with miRNA microarray (Figure 7.22, Figure 7.23).

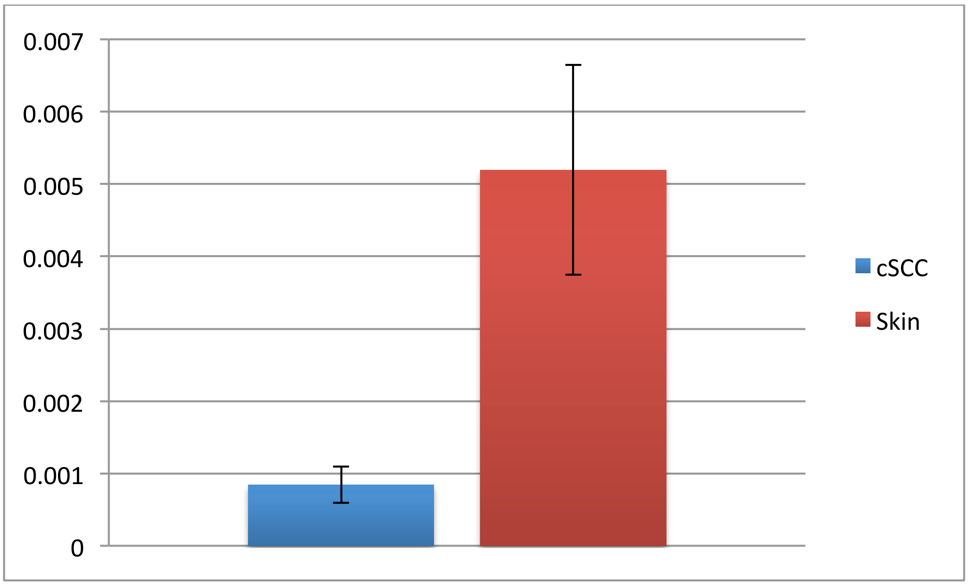


Figure 7.22 Validation of differential expression of has-miR-1908 in skin and cSCC. Bar chart showing 2^deltaCt values of hsa-miR-1908 expression in cSCC and skin samples. Down-regulation is concordant with changes detected with miRNA microarray. Whiskers=SEM.

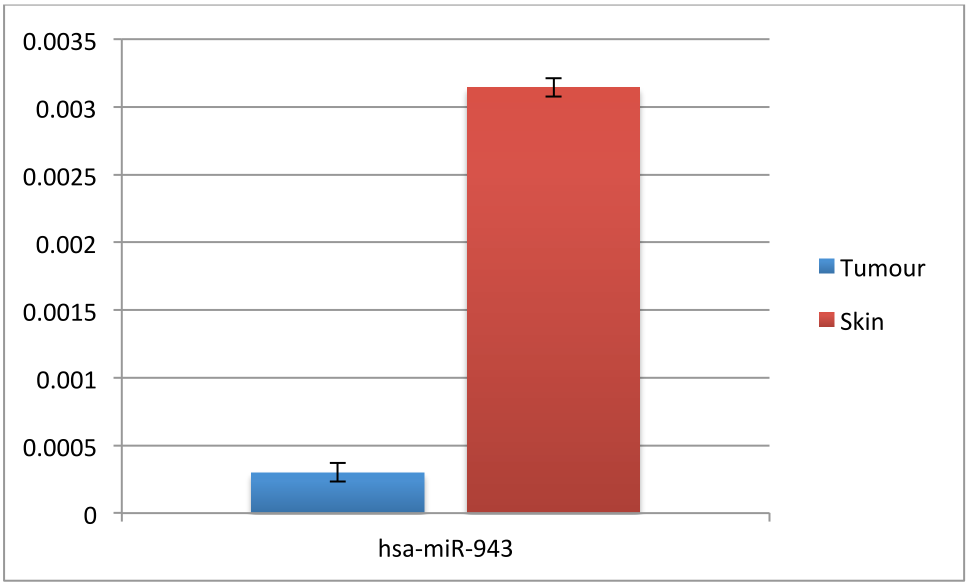


Figure 7.23 Validation of differential expression of has-miR-943 in skin and cSCC. Bar chart showing 2^deltaCt values of hsa-miR-943 expression in cSCC and skin samples. Down-regulation is concordant with changes detected with miRNA microarray. Whiskers=SEM.

7.4. Discussion

In this study, I explored the differences in miRNA transcription in cSCC and adjacent clinically normal skin. In comparison with previous studies, this investigation has utilised highly sensitive and specific microarrays that contain probes for over 2000 miRNAs and to date, the largest panel of clinical samples.

7.4.1. Global miRNA profile is sufficient to separate clinically normal skin from cSCC

Pearson's correlation of the global miRNA profile across all samples did not lead to complete separation of cSCC and adjacent normal skin samples, yet histologically equivalent samples collected from individual patients clustered together, suggesting a considerable influence of individual genetic background in miRNA profiling. Only 1 pair of skin and cSCC samples clustered together, which may be explained by proximity of those two clinical samples during collection, as the adjacent normal skin sample was usually collected from a surrounding rim of clinically normal skin during tumour removal. This may also explain findings in the PCA analysis in which several cSCC and normal skin samples clustered together. An alternative explanation includes technical issues during microarray processing (such as cross-contamination, environmental factors) that were not apparent during data processing. In addition to possible technical issues, the influence of low RIN value in several samples was considered as potential driver of this observation, yet RIN is likely not a critical factor, as most samples with low RIN clustered appropriately. Moreover, previous data have shown that miRNAs are highly stable even in clinical samples yielding RNA of low RIN (Jung et al., 2010).

Additionally, clinically normal skin adjacent to malignant tissue may harbour genetic mutations due to field cancerisation, in spite of clinically normal appearance (Kanjilal et al., 1995).

7.4.2. Several miRNAs are significantly differentially expressed in cSCC compared to normal skin

Our data show that miRNAs are not differentially transcribed in comparison between males and females, and I detected no statistically significant differences between immunosuppressed and immunocompetent patients.

Comparison of global miRNA profile of cSCC and adjacent normal skin, on the other hand, has detected 38 miRNAs with a p value <0.01, and 76 with a p value <0.05 as differentially expressed. Many of those miRNAs were previously implicated in cancer, including miR-943 shown to be up-regulated in HPV-18 positive Hela cells (Zhang et al., 2013), hsa-miR-744-5p found to be deregulated in head and neck cancer (Nurul-Syakima et al., 2011) and breast cancer (Enerly et al., 2011). MiR-642b-3p was previously shown to be up-regulated in pancreatic cancer (Ganepola et al., 2014), and hsa-miR-605 was previously linked with esophageal (Fassan et al., 2011) and gastrointestinal malignancies (Zhang et al., 2012).

The majority of differentially expressed miRNAs were down-regulated in cSCC. Down-regulation of miRNAs is common in cancer (Visone et al., 2007, Song et al., 2011), since miRNAs can function as both oncogenes and tumour suppressors. Additionally, miRNA genes are frequently located in areas of genomic alterations (Calin et al., 2004) and losses of genomic regions are more common in cSCC than gains, which may partially explain this finding.

Unsupervised hierarchical clustering of all samples based on expression levels of 38 top miRNAs has lead to almost complete segregation of cSCC and normal skin. 3 skin samples clustered with cSCC, one with substantial proximity to paired tumour sample (DT).

Additionally, this set of differentially expressed miRNAs has detected 9 miRNAs previously implicated in cSCC. Although we did not detect all miRNAs previously described to be dysregulated in cSCC, this may be explained by differences in methodology (using updated miRNA microarray for transcriptional profiling and using paired samples), and namely the high number of samples utilised in this study. The highest number of paired samples previously used for comprehensive miRNA exploration in cSCC was seven specimens; thereby this study represents the most comprehensive miRNA evaluation in cSCC to date.

7.4.3. Majority of differentially expressed miRNAs are concordant with genetic changes previously described in cSCC

Ten of sixteen miRNAs with known genomic location detected in our dataset are concordant with genomic alterations previously reported in cSCC (Purdie et al., 2009). A high frequency of genomic alterations of loci containing miRNA genes was previously reported in a variety of human tumours (Zhang et al., 2006a). The fact that the majority of differentially expressed miRNAs (62.5%) with known genomic location in our dataset integrated with known loci of genomic changes in cSCC identified by SNP array analysis suggests that miRNAs may be to some extent regulated by genetics, and the subsequent impact of miRNAs on transcriptional regulation may in fact reflect primary genetic changes, rather than act as an entirely independent epigenetic regulatory mechanism in this tumour. Given that genetic changes are notoriously difficult to address therapeutically, this stresses the attractiveness of miRNAs as potential clinical targets and warrants further exploration of their role in driving cSCC oncogenesis.

It is important to note that the genomic location of 20 (55%) differentially expressed miRNAs is not known, and may increase or decrease the proportion of miRNAs likely to be induced or downregulated due to genetic changes. Additionally, further regulatory processes and factors likely to play a role in miRNA dysregulation in cancer, since copy-number variation alone could not explain dysregulation of 37.5% of miRNAs with known genomic location.

7.4.4. Several viral miRNAs were found among miRNAs dysregulated in cSCC

Viral miRNAs were not previously detected in cSCC. Although the role of viral miRNAs in cancer has not been elucidated to date, it has been hypothesised that these may play a role as direct oncogenes, or may have a role in immunomodulation and cell cycle regulation (Pfeffer and Voinnet, 2006). Our dataset contains viruses from herpes simplex virus, Kaposi's Sarcoma-Associated Herpesvirus and cytomegalovirus, viruses which have all previously been implicated in cSCC (Zafiropoulos et al., 2003, Claudy et al., 1989). We found that 7/8 viral miRNAs were down-regulated in cSCC in comparison with normal skin. This may be explained by episomal latency of viruses, or more likely by their integration to genomic regions that is subsequently lost or gained due to copy-number variation in cSCC. An alternative explanation could be contamination of our samples with viruses, either during sample collection or during sample and mRNA processing, yet this explanation is less likely since the miRNAs are not amplified prior hybridisation to the array, and thus the minimal amounts of contaminating miRNAs are unlikely to lead to statistically significant differences. Our findings warrant further validation and exploration of viral miRNAs in cSCC development and progression.

7.4.5. Integration of *in silico* target prediction algorithm and transcriptional profiling of cSCC reveals novel miRNA targets in this malignancy

Integration of miRNA targets identified with Targetscan and transcription microarray data lead to the identification of 18 transcriptional targets. Among those targets, two genes, PDZD2, NRP2 were shown to be hypomethylated in cSCC and thereby theoretically of increased expression, yet PDZD2 is under-expressed at the transcriptional level, which suggests that this gene is possibly down-regulated by hsa-miR-1290. PDZD2 is a putative tumour suppressor found to be frequently hypermethylated in myeloid leukemia (Figueroa et al., 2010b), and neuropilin-2 is a member of neuropilin transmembrane glycoprotein family and a *bona fide* oncogene that functions as a receptor for VEGF binding and promotes cell migration and survival (Favier et al., 2006). Its upregulation in cSCC may be due to promoter hypomethylation, but our data shown in the previous chapter indicate that hypomethylation does not necessarily equate to an increase in transcription, thereby its upregulation may also be due to miRNA dysregulation, or a combination of both mechanisms.

Additional targets known to be involved in cancer were identified as potential miRNA targets. TGFBR3, downregulated in cSCC, is a member of TGF-beta signaling pathway that plays a critical role in tumourigenesis by interacting with MAPK p38α/β (Bragado et al., 2013) and its down-regulation has been linked with immunoevasion in melanoma and breast cancer (Hanks et al., 2013). CYP27B1 is a 25-hydroxyvitamin D3 1-alpha-hydroxylase that catalyses the conversion of calcidiol to the active form for vitamin D (calcitriol). Its increased expression has been described in a variety of cancers including colon cancer and breast cancer, and it has been hypothesised that local conversion of calcidiol into calcitriol may be responsible for tumour chemoresistance (Deeb et al., 2007). Ubiquitin-conjugating enzyme E2 H (UBE2H) targets proteins for degradation, and has been identified as a putative oncogene in hepatocellular carcinoma (Rajendra et al., 2004). Nedd4-binding partner 1 (N4BP1) regulates the function of Itch, which then ubiquitylates c-JUN and p63 (Oberst et al., 2007). Upregulation of N4BP1 inhibits the degradation of p63 through this mechanism (Graziano and De Laurenzi, 2011) and promotes tumourigenesis.

ETS2 is a transcription factor that has been shown to regulate p53 (Venanzoni et al., 1996) and has been shown to interact with MAPK pathway (Coffer et al., 1994). MAPKAP1 is another member of MAPK and also mTOR/AKT signalling pathway and has been previously reported to promote epithelial-to-mesenchymal transition (Xu et al., 2013). ACTN1 is another transcriptional target unregulated in our dataset, and its inhibition has been shown to increase tumour-free survival in mice (Craig et al., 2008), suggesting this gene may possess oncogenic properties.

Given that the panel of samples used for transcriptional profiling of sun-exposed, non-sunexposed skin and a subset of AK was treated with RNALater limited the possibility of comparing transcriptional profile of cSCC and normal skin as such comparison is likely to detect false targets (Chowdary et al., 2006). In order to correlate the expression levels of miRNA targets detected by comparison of AK and cSCC (fresh-frozen samples), trends in expression were displayed by using averaged transcriptional data across samples. This approach has confirmed expected expression trends in normal skin for 17 of the targets, which suggests that the process of oncogenic dysregulation is initiated early on and maintained through the progression from AK to cSCC.

7.4.6. Validation with miRCURY LNA™ Universal RT microRNA PCR confirms 2 miRNAs as differentially expressed

We validated the microarray findings in 3 pairs hybridised to miRNA microarrays and additional two pairs of cSCCs and adjacent normal skin pairs with miRCURY LNA™ Universal RT microRNA PCR. The selection of potential miRNAs for validation was highly limited by primers available for this technology, and only 1 miRNA detected in our data with a p value <0.01 was included in the panel of targets. The majority of miRNAs evaluated by this technology failed the process of quality control, most likely due to technical issues, as target detection was not equal in replicates of the same sample (data not shown). Two miRNAs were validated using this approach, including hsa-miR-943 included in top 38 dysregulated miRNAs.

7.4.7. Additional study limitations

Limitations of the study include using microarrays for miRNA detection and expression profiling, which likely limited the number of miRNAs identified as dysregulated in cSCC and the number of transcriptional targets. Moreover, no miRNA targets were functionally validated. Additionally, a relatively low number of samples from females and IC individuals were included in the panel.

7.5. Conclusions and future directions

This study confirms dysregulation of miRNAs in cSCC, and identified several novel miRNAs as dysregulated in this malignancy. The set contained several miRNAs previously implicated in cSCC, in addition to novel miRNAs including several viral miRNAs. Genes for the majority of detected miRNAs are located in areas of genetic copy-number variation in cSCC and their dysregulation is concordant with those changes. Additionally, using an *in silico* target prediction method and previously generated transcriptional data, we identified 18 miRNA targets, many of which were previously implicated in cancer. Several of those genes (N4BP1, TGFBR3, PDZD2) represent attractive targets for further exploration in cSCC.

Future directions include re-validating selected targets using standard qRT-PCR, exploring the correlation of global miRNA expression profile with cSCC tumour subtype, and validating selected putative miRNA targets using siRNA interference in cSCC cell lines.

8. Viruses and epigenetic regulation in non-melanoma skin cancer: prevalence of Merkel Cell Polyomavirus and the impact of MCPyV and human papillomavirus on methylation in cSCC

8.1. Introduction

As described in Chapter 1, human papillomavirus (HPV) has long been implicated in the pathogenesis of cSCC. More recently, DNA polymavirus named Merkel Cell polyomavirus (MCPyV) has been detected in a series of Merkel cell carcinoma (MCC) samples (Feng et al., 2008), and represents the first example of DNA polyomavirus integrating into the host genome. High prevalence of the virus in MCC samples was subsequently confirmed by several independent studies, ranging from 24 % (Garneski et al., 2009b) to 100 % of tested MCC samples (Sastre-Garau et al., 2009).

Since its discovery, the presence of MCPyV has been described in various types of lesions and populations: geographical differences in prevalence of the virus in MCC clinical samples were reported between European, American and Australian MCC patients (Garneski et al., 2009b). Additionally, the virus was shown to be significantly more common among HIV-positive men compared to healthy controls (Wieland et al., 2011).

Conversely, the virus was not detected in proliferative skin disorders (Foulongne et al., 2008), extracutaneous melanoma (Giraud et al., 2008), lymphoid tissue of AIDS patients (Sharp et al., 2009), prostate cancer samples (Bluemn et al., 2009), or small lung cancer (Wetzels et al., 2009), which morphologically resembles MCC. Non-small lung cancer samples from a cohort of Japanese patients, on the other hand, contained the virus in 9/32 (28 %) lung SCCs, and 9/45 (20 %) of lung adenocarcinomas (Hashida et al., 2013). Recently, it was shown that MCPyV may in fact play a role in chronic lymphocytic leukaemia: patients with CLL are at increased risk of MCC (Koljonen et al., 2009), and the virus was detected in 19 CLL samples, 6 of which also possessed novel large T antigen mutations (Koljonen et al., 2009). Chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL) specimen, a type of non-Hodgkin lymphoma associated with impaired immune system, were also shown to be positive for the virus in 33 % of tested samples, while the virus was not detected in follicular lymphoma samples assessed in the same study (Teman et al., 2011). This indicates that the virus may be associated with tumours linked with impaired immune system, and such immunosuppression may lead to its reactivation.

In non-melanoma skin cancer, 3/24 (12.5 %) of basal cell carcinomas were initially reported as positive for the virus in a cohort of European patients (Becker et al., 2009), with higher prevalence (72 %) subsequently reported among immunocompromised patients (Kassem et al., 2009). Yet a subsequent study of 11 BCC reported no positive samples (Andres et al., 2010) and this was confirmed in a series of 10 BCC collected from Japanese patients (Murakami et al., 2011), while another study of 41 BCC reported a prevalence of 29 % (Scola et al., 2012).

Reported positivity for MCPyV in cSCC ranged from 13 % (Garneski et al., 2009a) to 52 % in cSCC collected from immuncompromised individuals (Kassem et al., 2009). 4/30 (13 %) of cSCC from Japanese patients were positive for the virus (Murakami et al., 2011), while another study of 60 cSCC found 27 % positive samples (Scola et al., 2012). All these studies were carried out with FFPE samples. A large series of 145 fresh-frozen cSCC detected positivity in 38 % of samples, and showed a correlation between MCPyV positivity and serum antibodies to the virus (Rollison et al., 2012).

I aimed to verify the high prevalence of MCPyV in a series of MCC samples identified in the Barts and the London School of Medicine Pathology Group archive. I hypothesised that there may be significant differences in MCPyV positivity in non-melanoma skin cancer in transplant patients. Additionally, I hypothesised that there may be a correlation between HPV positivity and MCPyV presence in non-melanoma skin cancer samples. In broad terms, viral contributions to carcinogenesis may be regarded as epigenetic, thus I aimed to explore the impact of MCPyV positivity and HPV on methylation profile in CIS and cSCC samples.

8.2. Materials and methods

Formalin-fixed paraffin embedded samples of MCC collected from MCC patients, in addition to other benign, premalignant and malignant skin specimen from this patient cohort were obtained from the Pathology archive in collaboration with Pairaw Kader. Additional archival samples from MCC patients including blood were obtained through searching the experimental records of previous studies and the non-melanoma skin cancer archive at the Blizard institute. Table 8.1 lists all tissues and samples of MCC patients that were available, in addition to their basic clinical characteristics: a total of 8 MCCs, 1 blood sample and 13 non-MCC lesions were available in the MCC cohort. Histological sections of all tissues were reviewed and compared to the pathology reports to verify histological description of the tissues (Figure 8.1, Figure 8.2).

Patient	**Gender**	**Age**	**Immune status**	**Lesion collected**	**Location**
ML	F	77	IC	AK 1	-
				AK 2	-
				MCC primary	Head and neck
				MCC recurrent	Head and neck
				MCC nodal **metastases**	Head and neck

GC	F	86	IC	AK	-
				SCC	-
				Primary MCC	Trunk
				Carcinoma in situ	-
				Scar tissue	Trunk
EN	F	78	IC	Blood	-
				MCC 1 (incisional biopsy – FFPE)	Lower limb
				MCC 2 primary (FF)	Lower limb
				MCC 2 **metastases** (FF)	-
AD	M	80	IC	MCC primary	Upper limb
NR	F	76	IC	SCC	Unknown
				MCC primary	Lower limb
				MCC recurrent	Lower limb
JH	F	74	IC	MCC	Head and neck
				Granulation tissue	Head and neck
RW	M	70	RTR	SCC 1 (FF)	-
				SCC 2 (FF)	-
				SCC 3	-
				SCC 4	-
				AK	-
				BCC 1	-
				BCC 2	-
				MCC 1 (FF)	Head and neck
				MCC 1 nodal **metastases**	Neck lymph nodes
SH	F	-	IC	MCC primary	Upper arm

Table 8.1 Clinical characteristics of MCC patients included in the study and additional non-melanoma cancers and skin lesions collected from this patient cohort. FF=fresh frozen.

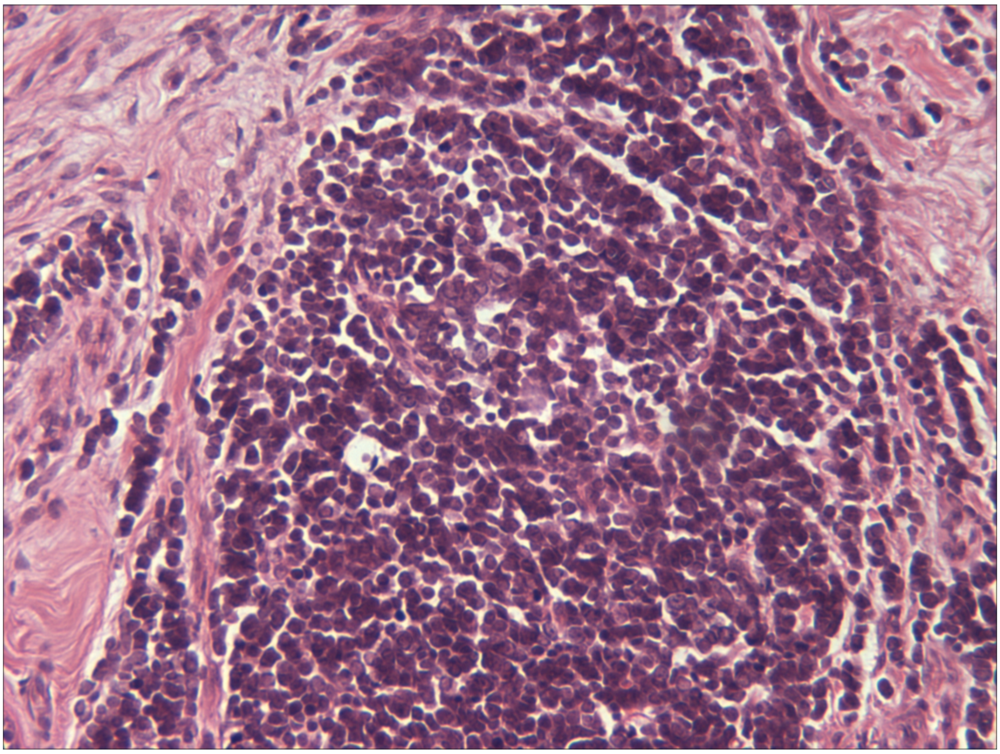


Figure 8.1 Haematoxylin and eosin staining of MCC (10x magnification) from a patient included in this series. This histological section illustrates small basophilic cells that highly resemble small cell lung carcinoma.

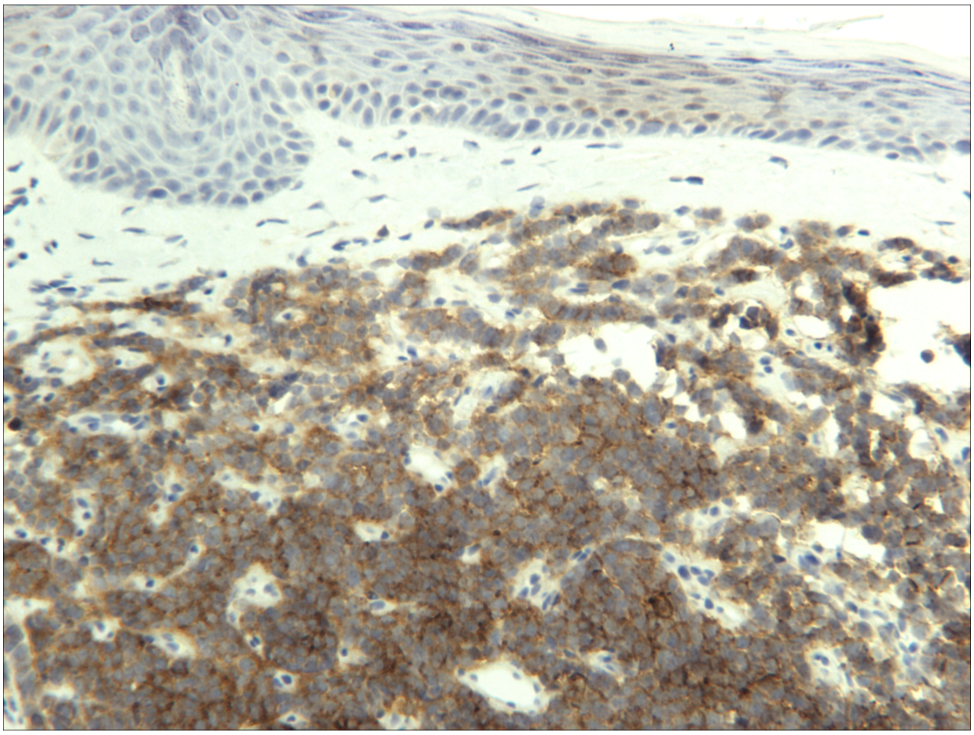


Figure 8.2 Merkel cell carcinoma CD56 positive staining. CD56 positivity is a typical immunohistochemical feature of MCC.

Additional set of 146 archival non-melanoma skin cancer DNA samples previously profiled for HPV virus presence was included in the study. This set contained a total of 42 BCC (29 from IC, 13 from OTR), 6 AK and 84 carcinoma in situ/cSCC (32 from immunocompetent patients, 58 from OTR), and 25 viral warts samples (all from OTR patients).

Detailed methodology including the process of DNA isolation from both fresh-frozen, FFPE tissues and DNA quantification is described in Chapter 2. Quality of all DNA samples was tested by PCR amplification of beta-globin gene prior to assessment of MCPyV presence. The presence of MCPyV was determined by a series of primers that amplify different regions of the virus: MCVPS1, VP1, LT1, LT3. All samples were assessed with a minimum of two MCPyV primers. HPV profiling was previously conducted on all archival samples using previously published method (Purdie et al., 2009).

8.2.1. Methylation profiling of archival cSCC samples

In order to determine the impact of MCPyV on DNA methylation in CIS/cSCC, 13 MCPyV positive and 15 negative (as detected by PCR amplification) archival cSCC DNA samples were hybridised to Infinium HumanMethylation450 array (described in greater details in the previous chapter). Table 8.2 provides clinical characteristics including date of collection and histological classification of all samples that passed the quality control and were included in the study. Samples hybridised in duplicate were labelled as _1 and _2 with corresponding sample code. DNA processing and hybridisation to the methylation array was carried out in collaboration with Dr Emily Clemente[17]. Data analysis was carried out by the author.

All samples were assessed for quality prior to normalisation, and all probes with known SNP and those corresponding to allosomes were removed from the dataset. Beta-values corresponding to the remaining probes were normalised using quantile normalisation in R, and statistical inference was determined by Welch t-test with p value adjustment using Bonferroni-Hochberg method.

ARRAY	SAMPLE	IMMUNE STATUS	TUMOUR SITE	HISTOLOGY	MCPYV	HPV	Date of collection
K16	K16	RTR	L shin/calf	SCC WD	-	HPV5	05/03/03
K17	K17	IC	R. Hand	SCC	-	HPV5	16/01/02
K18	K18	IC	R. Cheek	SCC WD	+	neg	03/04/02
K20	K20	ICP	L.Ear	SCC	+	DL347	01/11/02
K25	K25	RTR	R hand	SCC	-	HPV5, 14, 37	23/04/03
K26_1	K26	RTR	L. Hand	SCC	-	neg	25/06/03
K26_2	K26	RTR	L. Hand	SCC	-	neg	25/06/03
K30_1	K30	CT	Chest	SCC	+	neg	05/02/03
K30_2	K30	CT	Chest	SCC	+	neg	05/02/03

---

[17] Cambridge genomic Services, Department of Pathology, Cambridge University, Cambridge.

K38_1	K38	RTR	L. Shoulder	SCC	+	HPV19	07/01/98
K38_2	K38	RTR	L. Shoulder	SCC	+	HPV19	07/01/98
K39	K39	RTR	Dorsum R. Hand	SCC	+	neg	28/01/98
K40_1	K40	RTR	R. Shoulder	SCC	+	neg	26/11/03
K40_2	K40	RTR	R. Shoulder	SCC	+	neg	26/11/03
K42	K42	RTR	L. Forehead	SCC	+	HPV19	10/12/97
K43_1	K43	RTR	L. Ear	SCC	+	HPV24	29/01/97
K43_2	K43	RTR	L. Ear	SCC	-	HPV24	29/01/97
K46	K46	RTR	Dorsum L.Hand	SCC	-	neg	24/07/96
K47	K47	RTR	0	SCC	+	novel	20/09/95
K49	K49	RTR	L. Forearm	SCC	+	neg	09/07/96
K51_1	K51	IC	Forehead	SCC	+	HPV5	08/03/99
K51_2	K51	IC	Forehead	SCC	+	HPV5	08/03/99
K52	K52	RTR	Scalp	SCC	-	novel	14/01/99
K60	K60	RTR	R. Ear	CIS	-	neg	22/11/00
K63	K63	RTR	Temple	CIS	-	neg	03/01/01
K66_1	K66	ICP	L. Forehead	Bowens	-	RTRX7	12/03/01
K66_2	K66	ICP	L. Forehead	Bowens	-	RTRX7	12/03/01
K67	K67	RTR	R.ear	CIS	-	novel	04/04/01
K68	K68	RTR	R. hand	CIS	-	novel	04/04/01
K76	K76	RTR	L. Forearm	Bowens	-	HPV 5, 20	06/06/01
K82_1	K82	IC	L wrist	Bowens	-	Z95963, VS20-4	28/09/01
K82_2	K82	IC	L wrist	Bowens	-	Z95963, VS20-4	28/09/01
2H	2H	RTR	Hand	SCC	+	27	19/02/92
6R	6R	ICP	Ear	SCC	-	-	
6T	6T	ICP	Ear	SCC	+	-	
7E	7E	RTR	Hand	SCC	-	10,16,36,12r	

Table 8.2 Clinical characteristics of MCPyV positive and negative NMSC samples hybridised to Illumina 450K array. Bowens=Carcinoma in situ (CIS).

Comparison levels of methylation profile included MCPyV positivity, immune status, HPV positivity and CIS versus cSCC histology.

8.3. Results

8.3.1. Five of eight MCC samples were positive for MCPyV

PCR analysis of 8 MCC tumours detected MCPyV in 5/8 (62.5 %) of tested samples. Mean age at diagnosis of patients with MCPyV was 73.4 years (excluding one patient of unknown age at diagnosis), MCPyV-negative patients were on average 81.3 years old (p value=0.16).

All three metastatic MCC were MCPyV positive, and 2/3 nodal metastases of these tumours also showed positivity for the virus. One recurrent MCC of a MCPyV-positive primary tumour was negative for the virus, while the metastatic lesion was MCPyV-positive (patient ML).

Two patients with MCPyV-positive MCC and additional NMSC lesions (RW, ML) had detectable virus in at at least one additional lesion, while two patients with MCPyV-negative MCC and additional lesions (NR, GC) had no detectable virus in any of those lesions tested. Table 8.3 provides comprehensive results of all samples from MCC patients.

Patient	Lesion collected	Location	MCPyV
ML	AK 1	-	**Positive**
	AK 2	-	Negative
	MCC primary	Head and neck	**Positive**
	MCC recurrent	Head and neck	Negative
	MCC nodal metastases	Head and neck	**Positive**
GC	AK	-	Negative
	SCC	-	Negative
	Primary MCC	Trunk	Negative
	Carcinoma in situ	-	Negative
	Scar tissue	Trunk	Negative
EN	Blood	-	Negative
	MCC 1 (incisional biopsy – FFPE)	Lower limb	Negative
	MCC 2 primary (FF)	Lower limb	**Positive**
	MCC 2 metastatic (FF)	-	**Positive**
AD	MCC primary	Upper limb	**Positive**
NR	SCC	Unknown	Negative
	MCC primary	Lower limb	Negative
	MCC recurrent	Lower limb	Negative
JH	MCC	Head and neck	Negative
	Granulation tissue	Head and neck	Negative
RW	SCC 1 (FF)	-	Negative
	SCC 2 (FF)	-	**Positive**
	SCC 3	-	Negative

	SCC 4	-	Negative
	AK	-	Negative
	BCC 1	-	Negative
	BCC 2	-	Negative
	MCC 1 (FF)	Head and neck	**Positive**
	MCC 1 nodal metastasis	Neck lymph nodes	Negative
SH	MCC primary	Upper arm	**Positive**

Table 8.3 Results of MCPyV testing in a series of MCC samples and additional lesions from MCC patients.

8.3.2. 39 % of AK/CIS/cSCC lesions, 28.6 % of BCC samples and 56 % of viral warts are positive for MCPyV

8.3.2.1. MCPyV in AK/CIS/cSCC lesions

Of 90 AK/CIS/cSCC[18] (58 collected from transplant patients, 32 from immunocompetent individuals), 35 were positive for MCPyV (39 %). When stratified by immune status, 23/58 (39 %) and 12/32 (37 %) of these lesions were positive for the virus (p value=0.8, chi square).

8.3.2.2. MCPyV in BCC and viral warts

A total of 12/42 (28.6 %) BCC samples were positive for MCPyV. This was not statistically significantly different compared with cSCC-related lesions (p value=0.07, chi square). Stratification by immune status revealed 5/13 (38 %) BCC from immunosuppressed patients and 7/29 (24 %) of BCC from immunocompetent individuals as positive for the virus, which is not a statistically significant difference (p value=0.2, chi square).

Viral warts (all collected from transplant patients), were positive for MCPyV in 14/25 samples (56 %). This was statistically significantly higher prevalence compared with BCC (p value=0.02, chi square) even after adjustment for immune status, while not statistically significantly different in comparison with cSCC-related lesions (p value=0.1, chi square). Figure 8.3 depicts cumulative percentage of positive samples across all groups.

Linear regression did not detect a significant association between HPV positivity and MCPyV presence.

---

[18] grouped together for the purpose of this analysis as cSCC-related lesions

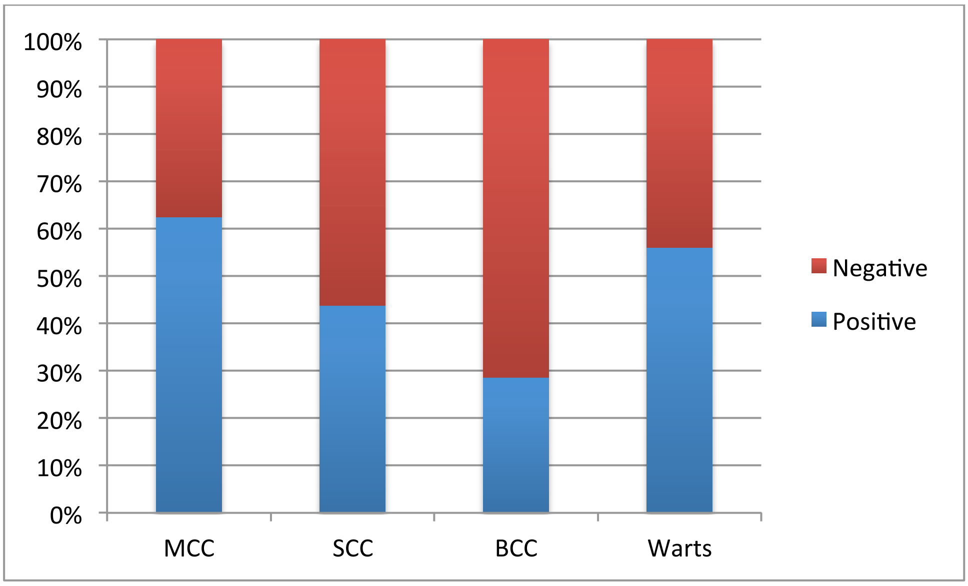


Figure 8.3 Cumulative percentage plot of MCPyV-positive NMSC samples. MCC samples showed the highest prevalence of the virus, followed by viral warts, cSCC and BCC lesions.

8.3.3. Quality control of cSCC archival samples hybridised to Illumina 450K methylation array

Archival DNA from 5 CIS and 23 cSCC samples was hybridised to Infinium HumanMethylation450 array (450K array). 8 samples were hybridised in duplicate to the array.

As shown in Figure 8.4, many samples showed poor quality based on beta-values distribution, and a total of 16 samples were subsequently excluded prior to normalisation, leaving 20 samples for analysis. Density plot of this sample set is shown in Figure 8.5. Not accounting for duplicate samples, 11 of those samples were negative for MCPyV, 9 were positive, 10 were positive for HPV and 10 were negative for HPV, 8 were CIS and 12 were cSCCs, 7 were from immunocompetent and 13 from immunosuppressed patients.

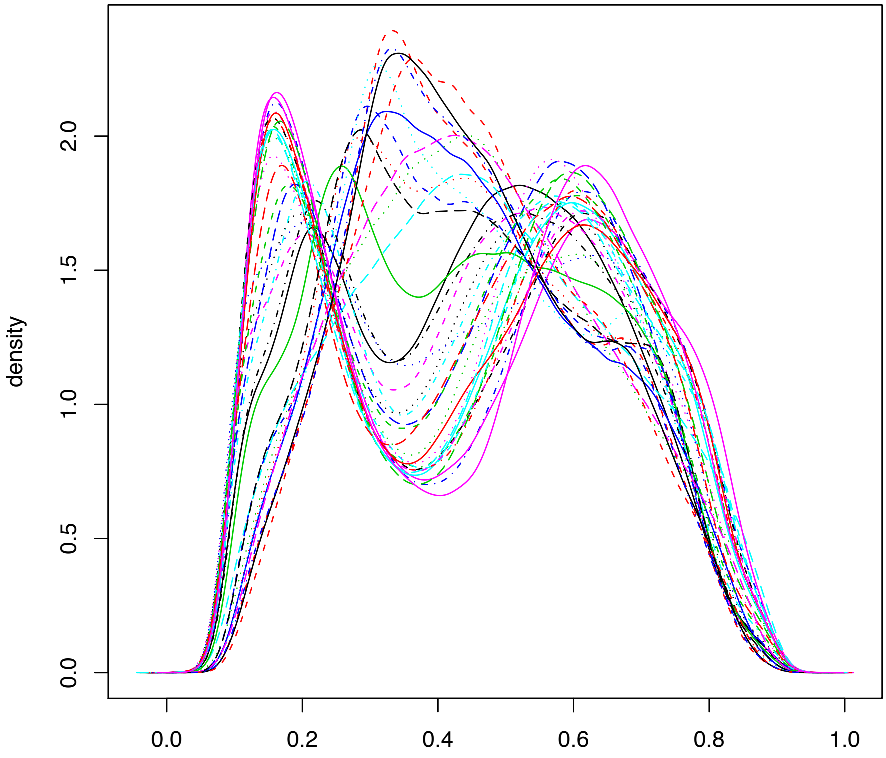


Figure 8.4 Density plot of beta-values of archival CIS/cSCC samples hybridised to Illumina 450K array. This figure shows that several samples do not follow expected pattern of bimodal distribution, and these samples were excluded from the dataset prior to normalisation.

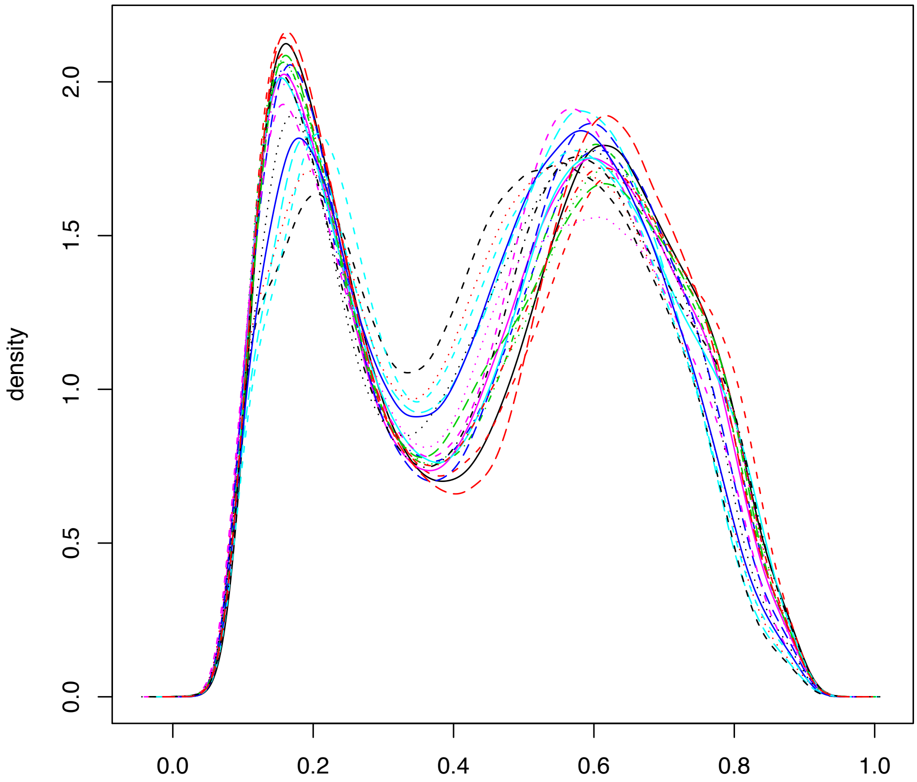


Figure 8.5 Density plot of archival CIS/cSCC samples that were included in the analysis.

8.3.4. Methylation profiling of CIS and cSCC archival samples using Illumina 450K methylation array reveals no differences in methylation due to MCPyV presence

Comparison of MCPyV-positive samples and MCPyV-negative CIS/cSCC samples revealed no differences in methylation between the two groups, and neither did comparison between HPV-positive or negative samples, between CIS and cSCC. Borderline significant differential methylation in one gene (p value=0.09), ZNF211 was detected by comparison of CIS/cSCC from transplant and immunocompetent patients.

8.4. Discussion and future directions

This study has found high prevalence of MCPyV in MCC samples, which is concordant with previous studies. Additionally, this is the first study that aimed to systematically identify differences in methylation between MCPyV positive CIS/cSCC and I detected differnces in MCPyV prevalence in a series of archival non-melanoma skin cancer lesions, and detected no correlation between HPV and MCPyV presence. Finally, by applying genome-wide methylation microarray profiling of archival CIS/cSCC samples, I detected no statistically significant differences in methylation in MCPyV or HPV positive and negative samples.

8.4.1. MCPyV-positive MCC patients are generally younger and all metastatic MCC were MCPyV positive

Our analysis detected expectedly high prevalence of MCPyV in MCC. Patients with the virus were on average almost a decade younger compared to their counterparts with MCPyV-negative lesions, and all MCC metastases were in MCPyV-positive patients (2/3 metastases were positive for the virus). Given that MCPyV has been shown to integrate in the genome, and although we did not confirm integration of the virus, this may indicate that its integration leads to increased genomic instability and generates a more aggressive MCC in younger patients. The clinical significance of MCPyV presence is widely discussed and significant advances have been made in understanding the molecular mechanisms through which it contributes to MCC development. However, to date, there is no consensus with respect to its clinical role in the clinical course of this tumour. Although our data are limited by the low number of samples and corresponding metastases, it warrants further exploration of the prognostic significance of the virus in MCC.

I also detected presence of the virus in at least one non-MCC lesion in patients with MCPyV-positive tumours, which indicates that there may be a systemic infection or skin-specific infection determining its presence in MCC and additional skin lesions of MCC patients.

8.4.2. MCPyV prevalence in NMSC is highest in viral warts among NMSC samples

I detected presence of the virus in 39 % of cSCC-related lesions, 28.6 % BCC and 56 % of viral warts. Such a high prevalence of the virus in viral warts represent a novel finding, and was statistically significant in comparison with the prevalence in BCC. Previously, only 1 viral wart of 16 samples was found to be positive for the virus using PCR (Mertz et al., 2010). Given that all viral warts samples were collected from transplant patients, I was not able to determine if such high prevalence is related to immunosuppression.

The prevalence of MCPyV detected in BCC and cSCC-related lesions represents a range that has been previously reported, but the absence of difference in prevalence between transplant and immunocompetent patients is in contrast with previously published data (Kassem et al., 2009), which reported prevalence of MCPyV in 52 % of cSCC and 72 % of BCC collected from transplant patients. This may be explicable by far higher number of CIS/AK/cSCC lesions in our study (35 versus 58), or by using fresh frozen DNA collected under strict sterile conditions in contrast with samples embedded in paraffin. I detected no correlation between MCPyV positivity and HPV presence.

8.4.3. Genome-wide methylation profiling of MCPyV- and HPV-positive and negative lesions revealed no differences in methylation between the two groups

This was the first study that attempted to identify potential regions of differential methylation in MCPyV- and HPV-positive NMSC samples.

Hybridisation of archival DNA to the Illumina450K array revealed not only major issues in quality of the produced data, which lead to the exclusion of 16 samples from the dataset, but also to a different distribution of beta-values in comparison with DNA freshly isolated from skin, AK and cSCC samples described in Chapter 6. While data from freshly isolated DNA show two peaks, one in a region of low methylation (0.0 to 0.2) and high methylation (0.8-1.0), data from archival samples shows peaks around 0.2 and 0.6, and practically no probes with either complete methylation or demethylation, which suggests that in addition to possible hybridisation issues, these seems to be loss of methylation and concentration of higher methylation values around 0.6. This could be due to poor hybridisation of degraded DNA (possibly due to repeated DNA freezing and thawing), or due to general loss of methylation due to long-term storage, although the methyl group is bound by a covalent bond.

This quality issue may explain the absence of significant findings between MCPyV- and HPV-positive and negative samples, in addition to the coverage of the array, since MCPyV may lead to differences in regions not assessed by the array.

8.4.4. Strengths and limitations of our study

One of the major strengths of our data is the inclusion of a large number of DNA isolated from fresh frozen NMSC lesions under strict sterile conditions which minimalised the risk of contamination with MCPyV through handling or paraffin embedding from both transplant and immunocompetent patients for which HPV profiling data also existed.

Limitations include a low number of MCC samples, and using only one technique to detect MCPyV, and not validating its prevalence by immunohistochemistry in additional set of samples.

8.4.5. Future directions

If a larger series of MCC samples could be collected, the presence of MPCyV could be evaluated as a prognostic factor in this malignancy.

The presence of MCPyV in additional series non-melanoma skin cancer samples (AK, CIS, BCC) and warts needs to be confirmed with immunohistochemistry with anti MCPyV large-T (LT) antigen antibody. Assessment of MCPyV integration and MCPyV viral load in non-melanoma skin cancer samples from both transplant and immunocompetent patients may reveal potential differences and provide further insights into the mechanisms by which samples the virus plays a possible oncogenic role.

Given that long-term storage of DNA samples can affect methylation levels, a definitive confirmation of no impact of MCPyV on methylation in non-melanoma skin cancer would involve profiling a large series of age-matched MCPyV-positive and negative samples with fresh DNA isolation using a whole-genome approach, since MCPyV may not affect genes or regions to which probes on a microarray correspond. A large number of such samples would be required, since the MCPyV is likely to affect a limited number of regions, which may also be entirely stochastic.

Additionally, correlation of MCPyV antibodies in patient blood and the presence of the virus in normal skin and NMSC samples and warts would show if a systemic infection with MCPyV predisposes patients to the development of these lesions.

9. Summary of results and concluding remarks

Epigenetic regulation has been extensively studied in a variety of human solid tumours, yet its detailed role in non-melanoma skin cancer has been largely unknown. Previous studies have implicated a limited number of differentially methylated genes and differentially expressed miRNAs as potentially relevant in non-melanoma skin cancer oncogenesis, but no comprehensive genome-wide assessment of epigenetic regulation of non-melanoma skin cancer has been carried out. In this work, I assessed genome-wide methylation in normal skin, actinic keratosis and cSCC tissues and cell lines, and integrated these data with gene expression and copy-number variation data in order to gain further insight into molecular regulation of this malignancy. Additionally, I explored miRNA regulation in cSCC and the impact of Merkel Cell Polyomavirus and human papillomavirus on DNA methylation in this tumour.

9.1.1. Genomic aberrations are significantly more common in AK compared to normal skin, and are largely stochastic in nature and have low correspondence with transcriptional changes

Copy-number variation analysis in normal skin and AK detected significantly more common genomic aberrations in AK compared to normal skin. Moreover, the size of regions of genomic imbalance was significantly greater compared to normal skin. This is a novel finding, and indicates that genetic changes underpin the onset of AK at least to some extent, although most of the genetic aberrations were largely stochastic and patient-specific, which may be partially explicable by individual dose and character of life-time sun-exposure. Regions of genomic changes overlapping in individual patients occurred mostly on chromosomes 8 and 9; chromosome 9 is a known hotspot for genetic LOH in AK. Genes potentially affected in AK across multiple patients included genes corresponding to amplification hotspots on chromosome 9 (SYK, TSC1) and known oncogenes (VAV2), but only two genes showed concordant genomic gain and increase in expression in AK: GAS1 and RXRA. The reason why so few genes showed expression change concordant with genomic aberrations may be due to the fact that a maximum of four AK contained a given region of genomic aberration and are thereby unlikely to propagate across all samples evaluated by gene transcription array. Additionally, AK samples in the transcription dataset were not entirely matched with samples for which SNP array data were available. Comparison of matched-only transcriptional and SNP array data would not be statistically appropriate due to the very low number of samples. Comparison of transcriptional and SNP data on an individual patient level would not provide information from which any statistical inference (and thereby construct validity) could be derived.

9.1.2. AK from male patients contain more genomic aberrations compared to females, and transplant patients more frequently show loss of genomic regions compared to immunocompetent counterparts

Additionally, this analysis revealed a statistically significant risk difference for genomic aberrations in AK in males compared to females (42%), and a significant difference between transplant and immunocompetent patients in the character of genomic imbalance: 40% of aberrations in transplant patients represented a loss of genomic material, while only 20% of SCNAs in immunocompetent patients represented a loss of a chromosomal region, which is a novel finding. This may be explicable by differences in the nature of genomic instability in immunosuppressed patients, or may be due to the study size.

9.1.3. Transcriptional profiling of normal skin and AK reveals dysregulation of genes involved in epidermal differentiation, cell-cell adhesions and WNT and PPAR pathways

Transcriptional profiling of non-sun exposed and sun-exposed skin and AK lesions detected 292 and 307 differentially expressed genes. A significant proportion of genes in both groups plays a role in epidermal development and cellular contact, the loss of which is a hallmark of malignant transformation and epithelial-to-mesenchymal transition. This finding indicates that processes that show a prominent role in cSCC (dedifferentiation, EMT) are initiated early on in the process of malignant transformation and are prominent in precancerous lesions. Therapeutic inhibition of this process in AK may thereby be of clinical relevance.

9.1.4. Transcriptional profiling of AK and cSCC reveals up-regulation of genes conferring invasive and metastatic potential in cSCC, and expression profiling correlates with histopathological classification

Comparison of AK and cSCC transcriptomes detected a total of 346 differentially expressed genes. Many highly up-regulated genes were shown to play a critical role in cancer invasion, including several metalloproteinases (MMP1, MMP3, MMP10), osteopontin (SPP1) and LAMC2. Additionally, NT5E (CD73) is highly up-regulated in cSCC, and its upregulation has been linked with increased metastatic potential in melanoma (Cui et al., 2012).

Transcriptional profiling of cSCC has also demonstrated significant transcriptional differences between various cSCC histological subtypes. This finding is largely consistent with copy-number variation profiling of cSCC samples, which found specific differences in genomic aberrations in poorly differentiated cSCC tumours (Purdie et al., 2009).

9.1.5. Increase in transcription of known stem cell markers heralds progression from AK to cSCC and distinguishes well differentiated tumours from poorly differentiated subtypes

Exploration of transcriptional levels of 41 known stem cell markers in AK and cSCC samples has shown distinct expression levels in AK, well-differentiated cSCC and poorly differentiated cSCC.

This finding indicates that expression of stem cell markers confers a certain degree of malignant potential in transformed cells, as evidenced by recent work implying SOX2 in mouse cSCC onset (Boumahdi et al., 2014) and NOTCH1 and NOTCH2 in human cSCC (Rehman et al., 1997). While this finding warrants further validation by additional techniques in an independent set of samples, comparison with transcriptional profiles of 4 independently generated mesenchymal stem cells types revealed higher correlation of poorly differentiated tumours with mesenchymal stem cells. Given that a subset of cSCC grow and progress rapidly (Rehman et al., 1994), evaluation of stem cell markers as biomarkers of cSCC invasiveness is warranted based on these observations.

9.1.6. Methylation profiling of cultured normal keratinocytes shows a high degree of correlation with normal skin, but only 21.3% of differentially methylated genes detected in cSCC cell lines were detected in cSCC clinical specimens

Cell culture is one of the most common laboratory techniques, and represents an attractive method for exploration of the methylome in skin due to the relative scarcity of normal skin samples. Comparison of short-term keratinocyte culture methylation data with methylation profiles of normal skin samples detected a high degree of correlation between the two tissues, suggesting that methylation is largely stable in short-tem keratinocyte culture.

The comparison of differentially methylated genes detected in cSCC cell lines with DMG detected in cSCC clinical samples revealed that only 21.3 % of genes detected in cSCC cell lines are differentially methylated in clinical specimens. Given that the number of normal skin (n=20) and cSCC samples (n=19) was far greater than the three cultured keratinocytes and 9 cSCC cell lines, it seems likely that statistical inference in cSCC cell lines was biased by the low number of control samples. On the other hand, all hyper- and hypomethylated genes validated in cSCC clinical samples showed concordant direction of change (hyper or hypomethylation), indicating that these genes are likely to represent genuine methylation targets in cSCC oncogenesis, rather than cell culture artifacts.

9.1.7. Genome-wide methylation profiling of normal skin, AK and cSCC revealed widespread gene body hypermethylation and hypomethylation of intergenic regions and malignant-transformed lesions

I carried out a systematic analysis of genome-wide methylation data in normal skin, AK and cSCC samples. I deteted no differences in methylation in normal skin due to sun-exposure, but found important differences in DNA methylation in AK and cSCC compared to normal skin. Our data also provide important observations regarding the functional genomic distribution of differentially methylated probes in AK and cSCC; we show that gene body regions are more frequently hypermethylated in both AK and cSCC, while intergenic regions are widely hypomethylated in those two tissues. The distribution of hyper- and hypomethylated promoters, on the other hand, was generally equal.

Given that previous studies of cSCC epigenetics have been focused largely on differences in promoter CpG methylation, this is an entirely novel finding that provides an insight into global epigenomic regulatory circuits in AK and cSCC. Gene body methylation has been previously described as evolutionarily conserved in both plants and higher organisms (Ibragimova et al., 2013), and thereby disruption in this pattern is likely to have major impact on the regulation of cellular functions. Integration of gene expression and DNA methylation data has revealed that gene body hypermethylation is associated with decreased gene transcription. This is contrary to previously reported findings, but may be explained by not stratifying gene body methylation by percentage, since the correlation between gene body methylation and gene expression has been previously reported as “parabolic” with a peak around 70% methylation followed by a sharp drop in the highest methylation ranges (Caruana, 2002). By averaging the the mean log2 rations, this subtle differentiation has been lost.

Intergenic hypomethylation been shown to be an important regulatory mark in stem cells (López-Bigas et al., 2002), and may confer a degree of regulatory plasticity in AK and cSCC, in addition to differential promoter and gene body methylation.

9.1.8. Differential methylation of CpG islands is more frequent in cSCC compared with AK, while Open Sea regions are frequently differentially methylation in AK

Over 50% of both hyper- and hypomethylated probes in cSCC compared to normal skin corresponded to CpG island and adjacent regions (Shores). In AK, on the other hand, CpG island probes represented 14% and 11% of hyper- and hypomethylated probes. This observation suggests that an important shift in positional methylation occurs in cSCC with a concentration of differences around CpG islands, while this convergence towards CpG-proximal regions is absent in AK. While this observation may be due to higher numbers of genomic aberrations in cSCC compared to AK, this may also represent a critical regulatory difference between premalignant and malignant skin lesions, with a strong potential for therapeutic targeting.

9.1.9. Differentially expressed miRNAs are predominantly down-regulated in cSCC compared to normal skin, and largely correspond to known regions of genomic alterations in cSCC

The comparison of miRNA microarray profiles of cSCC clinical samples and adjacent normal skin has revealed significant differences in the miRNA transcriptome of cSCC and normal skin, including differential expression of several miRNAs of viral origin, including herpes simplex virus, Kaposi's Sarcoma- Associated Herpesvirus and cytomegalovirus miRNAs. This observation is a novel finding in cSCC, and indicates a possible regulatory role of viruses in this malignancy.

Most differentially expressed miRNAs were down-regulated in cSCC compared to normal skin. I observed a high concordance of miRNA expression levels and known regions of chromosomal instability in cSCC. This indicates that miRNA dysregulation in cSCC may to some extent reflect genomic changes, rather than represent an entirely independent regulatory mechanism. Given the notorious difficulty of addressing genomic changes in human malignancies therapeutically, dysregulated miRNAs that convey malignant properties due to genomic aberrations may represent an attractive therapeutic target.

9.1.10. Merkel cell polyomavirus is common in non-melanoma skin cancer, but neither the presence of this virus nor human papillomavirus has a significant impact on methylation profiles of archival NMSC samples

I detected high prevalence of MCPyV in MCC samples in accordance with previous literature, and explored the presence of the virus in additional series of non-melanoma skin cancer samples. The prevalence of the virus detected across cSCC and BCC samples was largely concordant with the previous literature, while the unexpectedly high prevalence (56 %) of the virus in viral warts collected from organ transplant recipients represents a novel finding. While the absence of viral wart samples from immunocompetent patients limits the interpretation of this finding in the context of immunosuppression, the prevalence was statistically significantly higher compared to other lesions (BCC, cSCC) in transplant patients. Given that all viral warts in our dataset were positive for HPV, this finding may indicate a frequent “co-infection” of these two viruses in viral warts, or reactivation of MCPyV infection in transplant patients.

Genome-wide profiling of MCPyV- and HPV-positive and negative samples revealed no significant differences in methylation after p value adjustment. However, this analysis has detected important changes in the methylation profile of archival NMSC during quality control that has indicated substantial loss of methylation in archival DNA. Given this important technical aspect, in addition to the relatively low number of samples that passed the quality control and the bias of the Illumina Infinium HumanMethylation450 BeadChip towards certain regions of the genome, this finding requires further validation using freshly isolated DNA from cSCC samples using a whole-genome approach, such as bisulfite sequencing.

In summary, my work has shown that:

1. Genomic aberrations are significantly more common in AK compared to normal skin, and are also greater in size.

2. Regions of genomic imbalance in normal skin and actinic keratosis are largely stochastic and patient-specific, yet chromosome 8 and 9 contain focal SCNAs that occurred in multiple patients and represent potential hotspots for mutations in AK.
3. Transformation of normal skin to AK involves changes in genes relevant in cellular contact, epidermal differentiation and developmental pathways (Wnt pathway).
4. Transcriptional profiling of cSCC corresponds with histological subtyping.
5. Transcriptome profiling of poorly differentiated tumours is more highly correlated with that of mesenchymal stem cells, and well-differentiated tumours can be segregated from less differentiated tumours based on a set of known stem cell markers.
6. Gene bodies are largely hypermethylated, whilst intergenic regions are predominantly hypomethylated in AK and cSCC compared to normal skin.
7. Most areas of differential methylation in AK correspond to Open Sea regions, while differential methylation is largely concentrated in CpG islands and CpG-adjacent regions in cSCC.
8. Differentially expressed miRNAs in cSCC are largely down-regulated and many correspond to regions of known genomic alteration in cSCC.
9. Merkel cell polyomavirus is common in non-melanoma skin cancer and very common in viral warts.
10. Archival DNA is not suitable for detailed DNA methylation profiling due to DNA degradation.

Appendix 1. Differentially expressed genes between SE skin and AK.

Symbol	Chrom	logFC	adj.P.Val	Description
PHYHIP	8	2.231082407	3.07E-15	phytanoyl-CoA 2-hydroxylase interacting protein
WIF1	12	3.112869365	3.62E-15	WNT inhibitory factor 1
LEPR	1	2.278620413	3.62E-15	leptin receptor
NA	-	2.301669581	3.71E-15	NA
ADH1B	4	3.094469021	1.13E-11	alcohol dehydrogenase 1B (class I), beta polypeptide
SP8	7	2.683868401	3.09E-11	Sp8 transcription factor
KRT6B	12	-2.263015981	3.40E-11	keratin 6B
OASL	12	-3.210809595	5.83E-11	2'-5'-oligoadenylate synthetase-like
SYT17	16	1.881900085	7.77E-11	synaptotagmin XVII
BTC	4	2.253157435	1.31E-10	betacellulin
S100A9	1	-3.62786163	2.49E-10	S100 calcium binding protein A9
GJB2	13	-1.451842609	6.20E-10	gap junction protein, beta 2, 26kDa
KRT16	17	-3.317399133	8.77E-10	keratin 16
IGFL1	19	-3.862180153	9.24E-10	IGF-like family member 1
AGTR1	3	2.09561044	1.12E-09	angiotensin II receptor, type 1
CCL27	9	2.749677886	1.33E-09	chemokine (C-C motif) ligand 27
CYP4B1	1	2.327667074	1.77E-09	cytochrome P450, family 4, subfamily B, polypeptide 1
S100A8	1	-2.384712806	1.79E-09	S100 calcium binding protein A8
NETO2	16	-2.238239202	2.24E-09	neuropilin (NRP) and tolloid (TLL)-like 2
DSC2	18	-1.879176456	2.28E-09	desmocollin 2
HRASLS5	11	1.242229366	4.32E-09	HRAS-like suppressor family, member 5
CD24	6	-1.525641872	7.17E-09	CD24 molecule
ENAH	1	-1.963434489	8.31E-09	enabled homolog (Drosophila)
PCP4	21	2.692812549	8.40E-09	Purkinje cell protein 4
SLC7A5	16	-1.710791838	8.89E-09	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5
KRT9	17	-4.293852464	1.86E-08	keratin 9
PGM5	9	1.512594075	2.40E-08	phosphoglucomutase 5
C1QTNF7	4	1.590847144	2.50E-08	C1q and tumor necrosis factor related protein 7
RNF180	5	1.744334828	2.84E-08	ring finger protein 180

SOX5	12	1.762823808	2.84E-08	SRY (sex determining region Y)-box 5
RAB3B	1	1.087332509	3.62E-08	RAB3B, member RAS oncogene family
LOC572558	9	1.231956459	4.21E-08	hypothetical locus LOC572558
GALNT6	12	-1.52600612	4.21E-08	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)
PAMR1	11	1.807209462	4.69E-08	peptidase domain containing associated with muscle regeneration 1
C20orf24	20	-0.971755373	5.26E-08	chromosome 20 open reading frame 24
NP	14	-1.396470515	5.37E-08	nucleoside phosphorylase
SAMD9	7	-1.805519162	5.94E-08	sterile alpha motif domain containing 9
S100A2	1	-0.737778306	6.72E-08	S100 calcium binding protein A2
KIAA0101	15	-1.606246695	9.37E-08	KIAA0101
MAPK6	15	-1.007506201	1.10E-07	mitogen-activated protein kinase 6
SYNCRIP	6	-0.890391948	1.22E-07	synaptotagmin binding, cytoplasmic RNA interacting protein
FBLIM1	1	-1.267529986	1.58E-07	filamin binding LIM protein 1
CLEC7A	12	-1.939579658	2.00E-07	C-type lectin domain family 7, member A
MAPT	17	1.249277745	2.01E-07	microtubule-associated protein tau
MYO1B	2	-1.082698883	2.20E-07	myosin IB
WNT5A	3	-2.123208376	2.36E-07	wingless-type MMTV integration site family, member 5A
SORBS1	10	1.471180662	2.36E-07	sorbin and SH3 domain containing 1
TSPAN8	12	2.251691671	2.39E-07	tetraspanin 8
YARS	1	-0.69164725	3.01E-07	tyrosyl-tRNA synthetase
EZH1	17	0.712855518	3.02E-07	enhancer of zeste homolog 1 (Drosophila)
GSTM5	1	1.483811062	4.35E-07	glutathione S-transferase mu 5
SLC8A1	2	1.489506513	4.35E-07	solute carrier family 8 (sodium/calcium exchanger), member 1
SUSD2	22	1.692237418	4.35E-07	sushi domain containing 2
TXNDC17	17	-1.097147117	4.44E-07	thioredoxin domain containing 17
C4orf31	4	1.792256694	4.99E-07	chromosome 4 open reading frame 31
KCNJ15	21	-1.491503047	5.30E-07	potassium inwardly-rectifying channel, subfamily J, member 15

TNNC1	3	1.163279276	5.96E-07	troponin C type 1 (slow)
SYNE1	6	0.933865749	6.23E-07	spectrin repeat containing, nuclear envelope 1
CDK5R1	17	-1.462651258	6.69E-07	cyclin-dependent kinase 5, regulatory subunit 1 (p35)
KRT6A	12	-2.485070983	6.74E-07	keratin 6A
DEFB4	8	-4.660702841	6.95E-07	defensin, beta 4
GSTA3	6	1.893145666	7.02E-07	glutathione S-transferase alpha 3
RPL37	5	0.79024293	8.94E-07	ribosomal protein L37
KDM5A	12	0.763305142	9.02E-07	lysine (K)-specific demethylase 5A
LYVE1	11	1.616716411	1.05E-06	lymphatic vessel endothelial hyaluronan receptor 1
HSPB7	1	1.275963505	1.28E-06	heat shock 27kDa protein family, member 7 (cardiovascular)
RABIF	1	-0.962535046	1.31E-06	RAB interacting factor
PTTG1	5	-1.149935326	1.37E-06	pituitary tumor-transforming 1
TNFRSF21	6	-1.785235262	1.57E-06	tumor necrosis factor receptor superfamily, member 21
FUT3	19	-1.541417705	1.57E-06	fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group)
PDK4	7	1.297892027	1.57E-06	pyruvate dehydrogenase kinase, isozyme 4
COMP	19	-2.106298131	1.57E-06	cartilage oligomeric matrix protein
CAB39L	13	0.98206153	1.59E-06	calcium binding protein 39-like
ELL2	5	-0.924017765	1.59E-06	elongation factor, RNA polymerase II, 2
IFI27	14	-1.735121388	1.67E-06	interferon, alpha-inducible protein 27
POLR3G	5	-1.530982615	1.67E-06	polymerase (RNA) III (DNA directed) polypeptide G (32kD)
PI3	20	-3.084617491	1.74E-06	peptidase inhibitor 3, skin-derived
SCARA5	8	2.081592857	1.83E-06	scavenger receptor class A, member 5 (putative)
GTPBP4	10	-0.724622232	1.94E-06	GTP binding protein 4
PPARD	6	-0.997285688	2.00E-06	peroxisome proliferator-activated receptor delta
GJB6	13	-1.638101615	2.04E-06	gap junction protein, beta 6, 30kDa
C10orf99	10	-3.221956107	2.04E-06	chromosome 10 open reading frame 99
HOXC6	12	1.345349571	2.04E-06	homeobox C6
HOXB5	17	0.732636673	2.27E-06	homeobox B5
SNX1	15	0.838312	2.46E-06	sorting nexin 1
OAS1	12	-1.538597217	2.48E-06	2',5'-oligoadenylate synthetase 1, 40/46kDa

CADM2	3	0.90329592	2.53E-06	cell adhesion molecule 2
MYH11	16	1.731951048	2.71E-06	myosin, heavy chain 11, smooth muscle
AP2S1	19	-1.0272323	2.71E-06	adaptor-related protein complex 2, sigma 1 subunit
PPP4R1	18	-0.787571214	2.74E-06	protein phosphatase 4, regulatory subunit 1
SPINK6	5	-4.203949729	2.88E-06	serine peptidase inhibitor, Kazal type 6
C2orf67	2	0.973109162	2.91E-06	chromosome 2 open reading frame 67
GLRX3	10	-0.648886304	3.07E-06	glutaredoxin 3
RGMB	5	0.889799698	3.09E-06	RGM domain family, member B
HNMT	2	0.740542384	3.45E-06	histamine N-methyltransferase
C7orf10	7	-1.172716968	3.49E-06	chromosome 7 open reading frame 10
MXD1	2	-1.047973839	3.50E-06	MAX dimerization protein 1
ECE2	3	-1.04688826	3.53E-06	endothelin converting enzyme 2
C12orf5	12	-1.192739366	3.55E-06	chromosome 12 open reading frame 5
GTF3C6	6	-0.516190952	3.58E-06	general transcription factor IIIC, polypeptide 6, alpha 35kDa
TNFSF9	19	-1.411748182	3.58E-06	tumor necrosis factor (ligand) superfamily, member 9
PTPN21	14	0.969044184	3.63E-06	protein tyrosine phosphatase, non-receptor type 21
CSRP2	12	-1.485645516	3.72E-06	cysteine and glycine-rich protein 2
IL1F9	2	-2.806087243	3.90E-06	interleukin 1 family, member 9
HN1	17	-0.912186278	4.24E-06	hematological and neurological expressed 1
SCN7A	2	1.685879275	4.48E-06	sodium channel, voltage-gated, type VII, alpha
C1orf135	1	-1.236369182	4.60E-06	chromosome 1 open reading frame 135
DLX2	2	-2.420829236	4.90E-06	distal-less homeobox 2
SP5	2	1.025261467	5.17E-06	Sp5 transcription factor
N4BP2L1	13	1.152328855	5.34E-06	NEDD4 binding protein 2-like 1
ATOH8	2	0.983378311	5.62E-06	atonal homolog 8 (Drosophila)
SEMA3E	7	1.009808824	5.66E-06	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3E
FAM65C	20	-1.523886424	5.79E-06	family with sequence similarity 65, member C
GSN	9	0.968897048	5.79E-06	gelsolin (amyloidosis, Finnish type)
TPBG	6	-1.037909728	5.83E-06	trophoblast glycoprotein
SLC2A1	1	-1.363766571	5.87E-06	solute carrier family 2 (facilitated glucose transporter), member 1

HIF3A	19	1.653193341	5.87E-06	hypoxia inducible factor 3, alpha subunit
ZNF677	19	1.588365263	6.12E-06	zinc finger protein 677
TMOD1	9	0.966159126	6.59E-06	tropomodulin 1
HSPA14	10	-0.627767157	6.84E-06	heat shock 70kDa protein 14
IRAK1	X	-0.61437479	6.90E-06	interleukin-1 receptor-associated kinase 1
CASQ2	1	1.551476299	7.09E-06	calsequestrin 2 (cardiac muscle)
SLC30A9	4	-0.985849954	7.40E-06	solute carrier family 30 (zinc transporter), member 9
GALNTL1	14	1.565978335	7.50E-06	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 1
DSG3	18	-1.21683121	7.71E-06	desmoglein 3 (pemphigus vulgaris antigen)
ACTG2	2	1.852662859	7.84E-06	actin, gamma 2, smooth muscle, enteric
RALA	7	-0.633711725	7.85E-06	v-ral simian leukemia viral oncogene homolog A (ras related)
LEPROTL1	8	-0.734055955	8.25E-06	leptin receptor overlapping transcript-like 1
SUB1	5	-0.689469551	8.38E-06	SUB1 homolog (S. cerevisiae)
LCE3D	1	-3.714293176	8.63E-06	late cornified envelope 3D
FBXO45	3	-1.134901057	8.63E-06	F-box protein 45
SPINK7	5	-3.461361864	8.86E-06	serine peptidase inhibitor, Kazal type 7 (putative)
EIF4E	4	-0.621215178	8.91E-06	eukaryotic translation initiation factor 4E
TMEM63A	1	0.718057727	9.10E-06	transmembrane protein 63A
KRT17	17	-1.570923009	9.33E-06	keratin 17
LOC643650	10	1.348797179	9.36E-06	hypothetical protein LOC643650
SGCG	13	1.397768707	9.36E-06	sarcoglycan, gamma (35kDa dystrophin-associated glycoprotein)
RAI2	X	1.249504217	9.36E-06	retinoic acid induced 2
DENND4C	9	0.716868104	9.68E-06	DENN/MADD domain containing 4C
TNFAIP1	17	-0.623608342	9.93E-06	tumor necrosis factor, alpha-induced protein 1 (endothelial)
IRF7	11	-0.848459149	9.93E-06	interferon regulatory factor 7
CLDN11	3	1.564202995	1.03E-05	claudin 11
ADAP2	17	-1.377588992	1.07E-05	ArfGAP with dual PH domains 2
LRRC2	3	1.118890303	1.20E-05	leucine rich repeat containing 2
ENO1	1	-0.681648517	1.22E-05	enolase 1, (alpha)
CRTAP	3	0.840667118	1.22E-05	cartilage associated protein
GBP6	1	-2.019403844	1.22E-05	guanylate binding protein family, member 6

TNMD	X	1.578061911	1.25E-05	tenomodulin
FOSL1	11	-1.795895435	1.27E-05	FOS-like antigen 1
EIF6	20	-0.696817359	1.30E-05	eukaryotic translation initiation factor 6
ABCA8	17	1.43951639	1.30E-05	ATP-binding cassette, sub-family A (ABC1), member 8
KIAA1199	15	-0.646109238	1.31E-05	KIAA1199
SGSM1	22	0.576860422	1.31E-05	small G protein signaling modulator 1
ACVR2A	2	0.842663084	1.31E-05	activin A receptor, type IIA
MRPS14	1	0.697863157	1.33E-05	mitochondrial ribosomal protein S14
DLX1	2	-2.197751512	1.33E-05	distal-less homeobox 1
EIF5	14	-0.528228744	1.34E-05	eukaryotic translation initiation factor 5
FOXE1	9	-1.915008435	1.37E-05	forkhead box E1 (thyroid transcription factor 2)
MAMDC2	9	1.428412443	1.39E-05	MAM domain containing 2
AKR1B10	7	-2.490402458	1.41E-05	aldo-keto reductase family 1, member B10 (aldose reductase)
C12orf56	12	-1.764593231	1.42E-05	chromosome 12 open reading frame 56
VSNL1	2	-1.07087827	1.43E-05	visinin-like 1
FABP7	6	3.404045922	1.47E-05	fatty acid binding protein 7, brain
TK1	17	-0.822452932	1.47E-05	thymidine kinase 1, soluble
C14orf132	14	0.990892123	1.47E-05	chromosome 14 open reading frame 132
PLN	6	1.951613961	1.56E-05	phospholamban
GDA	9	-1.582776852	1.58E-05	guanine deaminase
CKS2	9	-1.093273628	1.59E-05	CDC28 protein kinase regulatory subunit 2
CMAH	6	1.132932633	1.66E-05	cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMP-N-acetylneuraminate monooxygenase) pseudogene
DCUN1D5	11	-0.617839525	1.66E-05	DCN1, defective in cullin neddylation 1, domain containing 5 (S. cerevisiae)
OGN	9	1.367433668	1.78E-05	osteoglycin
ERBB4	2	1.161623978	1.78E-05	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)
SRXN1	20	-1.061824985	1.79E-05	sulfiredoxin 1 homolog (S. cerevisiae)
FAM162A	3	-0.874927878	1.81E-05	family with sequence similarity 162, member A
LRRN4CL	11	1.418917357	1.85E-05	LRRN4 C-terminal like
FAM13A	4	0.795113822	1.86E-05	family with sequence similarity

				13, member A
ADAM23	2	-1.999949484	1.88E-05	ADAM metallopeptidase domain 23
EBNA1BP2	1	-0.508891279	1.93E-05	EBNA1 binding protein 2
ULBP2	6	-1.461372151	1.95E-05	UL16 binding protein 2
C1orf56	1	-0.615316791	2.05E-05	chromosome 1 open reading frame 56
GGH	8	-1.58208049	2.06E-05	gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)
UTP11L	1	-0.469248002	2.08E-05	UTP11-like, U3 small nucleolar ribonucleoprotein, (yeast)
BNC1	15	-0.906989724	2.11E-05	basonuclin 1
TMC8	17	-0.983228408	2.14E-05	transmembrane channel-like 8
MICALL1	22	-0.782810129	2.18E-05	MICAL-like 1
CALML3	10	-1.820669812	2.22E-05	calmodulin-like 3
FAM150B	2	1.321271756	2.29E-05	family with sequence similarity 150, member B
OAS2	12	-1.765115722	2.29E-05	2'-5'-oligoadenylate synthetase 2, 69/71kDa
CCRN4L	4	-1.459167504	2.31E-05	CCR4 carbon catabolite repression 4-like (S. cerevisiae)
N4BP2L2	13	0.871362327	2.33E-05	NEDD4 binding protein 2-like 2
NME1	17	-0.633154152	2.37E-05	non-metastatic cells 1, protein (NM23A) expressed in
GINS3	16	-0.636997509	2.40E-05	GINS complex subunit 3 (Psf3 homolog)
S100A16	1	-0.817695136	2.41E-05	S100 calcium binding protein A16
IFRD2	3	-0.760453901	2.41E-05	interferon-related developmental regulator 2
MPZL2	11	-1.430762183	2.48E-05	myelin protein zero-like 2
FILIP1	6	1.044050818	2.48E-05	filamin A interacting protein 1
PLA2G4D	15	-0.982822528	2.63E-05	phospholipase A2, group IVD (cytosolic)
GLTSCR2	19	0.602744463	2.83E-05	glioma tumor suppressor candidate region gene 2
UBE2F	2	-0.781502631	2.92E-05	ubiquitin-conjugating enzyme E2F (putative)
PSMD11	17	-0.553228174	2.95E-05	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11
FABP5	8	-0.986607334	2.96E-05	fatty acid binding protein 5 (psoriasis-associated)
PAR5	15	1.045987817	3.00E-05	Prader-Willi/Angelman syndrome-5
C11orf67	11	0.560151934	3.02E-05	chromosome 11 open reading frame 67
DDX52	17	-0.95323385	3.05E-05	DEAD (Asp-Glu-Ala-Asp) box

				polypeptide 52
OBFC1	10	-0.745207093	3.20E-05	oligonucleotide/oligosaccharide-binding fold containing 1
FGFBP2	4	1.753642781	3.41E-05	fibroblast growth factor binding protein 2
CDC42SE1	1	-0.492894535	3.45E-05	CDC42 small effector 1
ENSA	1	-0.813152838	3.46E-05	endosulfine alpha
ARL4C	2	-0.783047254	3.50E-05	ADP-ribosylation factor-like 4C
C7orf58	7	1.076736743	3.50E-05	chromosome 7 open reading frame 58
KIAA0802	18	-1.384517901	3.50E-05	KIAA0802
SFXN1	5	-0.553153344	3.53E-05	sideroflexin 1
LRRC20	10	-1.058692931	3.64E-05	leucine rich repeat containing 20
HK2	2	-0.930613232	3.65E-05	hexokinase 2
HOMER1	5	-0.963991288	3.70E-05	homer homolog 1 (Drosophila)
SMOX	20	-1.457804841	3.70E-05	spermine oxidase
CNN1	19	1.096422885	3.70E-05	calponin 1, basic, smooth muscle
LOC100130097	1	1.406252911	3.72E-05	hypothetical LOC100130097
RIMS3	1	-1.004627903	3.75E-05	regulating synaptic membrane exocytosis 3
WFS1	4	0.611474552	3.80E-05	Wolfram syndrome 1 (wolframin)
RAN	12	-0.579047554	3.82E-05	RAN, member RAS oncogene family
ZNF323	6	-1.043961478	3.82E-05	zinc finger protein 323
LEP	7	1.744600864	3.84E-05	leptin
FKBPL	6	-0.752997816	3.90E-05	FK506 binding protein like
NPY1R	4	1.138944059	3.92E-05	neuropeptide Y receptor Y1
ABCB6	2	-0.717186815	4.01E-05	ATP-binding cassette, sub-family B (MDR/TAP), member 6
BCL2A1	15	-1.380133147	4.02E-05	BCL2-related protein A1
SERPINB13	18	-2.194013705	4.02E-05	serpin peptidase inhibitor, clade B (ovalbumin), member 13
KANK2	19	0.847122796	4.05E-05	KN motif and ankyrin repeat domains 2
GSPT1	16	-0.4513105	4.05E-05	G1 to S phase transition 1
CNTNAP2	7	-2.257132986	4.05E-05	contactin associated protein-like 2
UST	6	1.216389885	4.05E-05	uronyl-2-sulfotransferase
C15orf59	15	1.074590895	4.12E-05	chromosome 15 open reading frame 59
SYNM	15	1.827257685	4.12E-05	synemin, intermediate filament protein
FHL1	X	1.168944927	4.13E-05	four and a half LIM domains 1
HIF1A	14	-0.886352709	4.32E-05	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)

RFTN1	3	-1.389136874	4.32E-05	raftlin, lipid raft linker 1
LDLR	19	-0.902604389	4.32E-05	low density lipoprotein receptor
RPL26L1	5	-0.615520382	4.34E-05	ribosomal protein L26-like 1
TMEM198	2	-0.754524134	4.35E-05	transmembrane protein 198
DIXDC1	11	0.993157805	4.72E-05	DIX domain containing 1
CLCA2	1	-0.857394247	4.73E-05	chloride channel accessory 2
FAM49B	8	-0.630998728	4.78E-05	family with sequence similarity 49, member B
NDRG4	16	-1.583361104	4.96E-05	NDRG family member 4
POU2F3	11	0.990707472	5.01E-05	POU class 2 homeobox 3
ALDH3A2	17	0.903345231	5.01E-05	aldehyde dehydrogenase 3 family, member A2
CABC1	1	0.749168853	5.05E-05	chaperone, ABC1 activity of bc1 complex homolog (S. pombe)
USP54	10	0.667402309	5.06E-05	ubiquitin specific peptidase 54
MARS	12	-0.532921537	5.07E-05	methionyl-tRNA synthetase
GSTP1	11	-0.769210503	5.11E-05	glutathione S-transferase pi 1
ID4	6	1.227801869	5.29E-05	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
CENPN	16	-1.051985247	5.36E-05	centromere protein N
CDC45L	22	-0.72038759	5.53E-05	CDC45 cell division cycle 45-like (S. cerevisiae)
SPRR1A	1	-2.310315099	5.53E-05	small proline-rich protein 1A
SYNPO2	4	1.383422642	5.53E-05	synaptopodin 2
SLC3A2	11	-0.712694004	5.59E-05	solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2
ZFYVE21	14	0.538800228	5.68E-05	zinc finger, FYVE domain containing 21
RNASEH1	2	-0.458794876	5.69E-05	ribonuclease H1
GPR1	2	-1.046417274	5.69E-05	G protein-coupled receptor 1
GSTM3	1	1.518028394	5.71E-05	glutathione S-transferase mu 3 (brain)
ISCA1	9	-0.592527822	5.90E-05	iron-sulfur cluster assembly 1 homolog (S. cerevisiae)
GPRIN1	5	-0.985742913	6.00E-05	G protein regulated inducer of neurite outgrowth 1
YWHAZ	8	-0.516212708	6.05E-05	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
SLAMF7	1	-1.630383541	6.20E-05	SLAM family member 7
SLC47A1	17	1.118856086	6.23E-05	solute carrier family 47, member 1
DLG2	11	1.293612966	6.26E-05	discs, large homolog 2 (Drosophila)
GDPD3	16	-1.489128494	6.26E-05	glycerophosphodiester

				phosphodiesterase domain containing 3
BBS2	16	0.703925117	6.28E-05	Bardet-Biedl syndrome 2
GPAM	10	1.012206427	6.41E-05	glycerol-3-phosphate acyltransferase, mitochondrial
SPRR3	1	-3.083151405	6.54E-05	small proline-rich protein 3
PRUNE2	9	1.495218852	6.56E-05	prune homolog 2 (Drosophila)
PSMD12	17	-0.558586985	6.60E-05	proteasome (prosome, macropain) 26S subunit, non-ATPase, 12
C6orf162	6	0.791423049	6.60E-05	chromosome 6 open reading frame 162
MYD88	3	-0.806136188	6.74E-05	myeloid differentiation primary response gene (88)
BCL2	18	0.633690207	6.76E-05	B-cell CLL/lymphoma 2
MBP	18	0.901487606	7.12E-05	myelin basic protein
KIAA1467	12	1.340519463	7.14E-05	KIAA1467
RCAN1	21	-1.064604596	7.18E-05	regulator of calcineurin 1
KPNA2	17	-0.723735253	7.42E-05	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)
PPIF	10	-0.980451637	7.50E-05	peptidylprolyl isomerase F
EEF2K	16	1.029665644	7.60E-05	eukaryotic elongation factor-2 kinase
AP1S3	2	-1.042497787	7.60E-05	adaptor-related protein complex 1, sigma 3 subunit
GNA15	19	-0.902996254	7.62E-05	guanine nucleotide binding protein (G protein), alpha 15 (Gq class)
DKK2	4	1.264256705	7.81E-05	dickkopf homolog 2 (Xenopus laevis)
SLC7A11	4	-1.663131251	7.81E-05	solute carrier family 7, (cationic amino acid transporter, y+ system) member 11
PRMT2	21	0.470093445	7.87E-05	protein arginine methyltransferase 2
ARPC5L	9	-0.67408019	8.13E-05	actin related protein 2/3 complex, subunit 5-like
TNPO1	5	-0.676331739	8.17E-05	transportin 1
UBE2T	1	-1.018716341	8.21E-05	ubiquitin-conjugating enzyme E2T (putative)
TNXB	6	0.787124512	8.25E-05	tenascin XB
DIDO1	20	0.530268909	8.25E-05	death inducer-obliterator 1
CDCA4	14	-0.630924316	8.50E-05	cell division cycle associated 4
F12	5	-1.240004245	8.62E-05	coagulation factor XII (Hageman factor)
KIAA1377	11	1.008937074	8.78E-05	KIAA1377
DUSP14	17	-1.200344537	9.19E-05	dual specificity phosphatase 14
C6orf153	6	-0.643117067	9.30E-05	chromosome 6 open reading frame 153
C11orf48	11	-0.526657615	9.33E-05	chromosome 11 open reading

				frame 48
SERPINB3	18	-2.843983856	9.37E-05	serpin peptidase inhibitor, clade B (ovalbumin), member 3
GPR177	1	0.691612027	9.38E-05	G protein-coupled receptor 177
ADH7	4	-2.00723802	9.42E-05	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide
PLBD1	12	-1.309521543	9.48E-05	phospholipase B domain containing 1
PLCB4	20	1.005686208	9.48E-05	phospholipase C, beta 4
FLNC	7	0.802395558	9.65E-05	filamin C, gamma
LGR5	12	1.10048201	9.65E-05	leucine-rich repeat-containing G protein-coupled receptor 5
H2AFY	5	-0.535907178	9.65E-05	H2A histone family, member Y
TMEM49	17	-1.213088689	9.66E-05	transmembrane protein 49
SEMA3D	7	0.925658085	9.72E-05	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D

Appendix 2. Full list of significantly overrepresented GO terms comparing AK and SE.

adjustedP	GO	Name	Probes
4.55E-39	GO:0001959	regulation of cytokine-mediated signaling pathway	7
1.32E-38	GO:0032597	B cell receptor transport into membrane raft	6
1.32E-38	GO:0032600	chemokine receptor transport out of membrane raft	6
1.32E-38	GO:0032913	negative regulation of transforming growth factor-beta3 production	6
1.55E-32	GO:0002768	immune response-regulating cell surface receptor signaling pathway	6
5.99E-32	GO:0004024	alcohol dehydrogenase activity, zinc-dependent	5
5.99E-32	GO:0033210	leptin-mediated signaling pathway	5
9.60E-29	GO:0042104	positive regulation of activated T cell proliferation	7
5.54E-28	GO:0001775	cell activation	6
4.85E-26	GO:0006069	ethanol oxidation	5
1.95E-24	GO:0030856	regulation of epithelial cell differentiation	6
1.95E-24	GO:0043408	regulation of MAPKKK cascade	6
9.12E-24	GO:0005432	calcium:sodium antiporter activity	5
2.20E-22	GO:0051924	regulation of calcium ion transport	7
8.12E-22	GO:0016500	protein-hormone receptor activity	5
1.33E-21	GO:0002237	response to molecule of bacterial origin	6
1.33E-21	GO:0031295	T cell costimulation	6
1.74E-20	GO:0043627	response to estrogen stimulus	13
3.99E-20	GO:0001542	ovulation from ovarian follicle	5
1.33E-19	GO:0031669	cellular response to nutrient levels	4
1.33E-19	GO:0051346	negative regulation of hydrolase activity	4
2.81E-19	GO:0030296	protein tyrosine kinase activator activity	6
2.81E-19	GO:0042325	regulation of phosphorylation	6
9.25E-19	GO:0007204	elevation of cytosolic calcium ion concentration	13
1.72E-16	GO:0045730	respiratory burst	6
8.31E-16	GO:0060401	cytosolic calcium ion transport	4
3.88E-15	GO:0006112	energy reserve metabolic process	5
5.84E-15	GO:0002026	regulation of the force of heart contraction	7
2.92E-13	GO:0005666	DNA-directed RNA polymerase III complex	3
2.92E-13	GO:0015276	ligand-gated ion channel activity	3
4.30E-13	GO:0002028	regulation of sodium ion transport	4
4.30E-13	GO:0015081	sodium ion transmembrane transporter activity	4
4.30E-13	GO:0043149	stress fiber assembly	4
4.98E-13	GO:0015758	glucose transport	7
1.39E-12	GO:0042755	eating behavior	5

3.35E-12	GO:0048706	embryonic skeletal system development	7
5.25E-12	GO:0008366	axon ensheathment	4
5.25E-12	GO:0017157	regulation of exocytosis	4
5.25E-12	GO:0048251	elastic fiber assembly	4
7.02E-12	GO:0035094	response to nicotine	5
9.03E-12	GO:0006936	muscle contraction	13
1.78E-11	GO:0001874	zymosan receptor activity	3
1.78E-11	GO:0001879	detection of yeast	3
1.78E-11	GO:0005859	muscle myosin complex	3
1.78E-11	GO:0018125	peptidyl-cysteine methylation	3
1.78E-11	GO:0043292	contractile fiber	3
4.81E-10	GO:0001846	opsonin binding	3
4.81E-10	GO:0002752	cell surface pattern recognition receptor signaling pathway	3
4.81E-10	GO:0005924	cell-substrate adherens junction	3
4.81E-10	GO:0006359	regulation of transcription from RNA polymerase III promoter	3
4.81E-10	GO:0009756	carbohydrate mediated signaling	3
4.81E-10	GO:0030241	skeletal muscle myosin thick filament assembly	3
4.81E-10	GO:0042287	MHC protein binding	3
5.94E-10	GO:0030054	cell junction	26
7.53E-10	GO:0043406	positive regulation of MAP kinase activity	6
5.14E-09	GO:0006006	glucose metabolic process	8
7.15E-09	GO:0021756	striatum development	3
7.15E-09	GO:0032982	myosin filament	3
1.15E-08	GO:0008544	epidermis development	11
3.35E-08	GO:0008629	induction of apoptosis by intracellular signals	6
3.40E-08	GO:0071109	superior temporal gyrus development	4
6.80E-08	GO:0002238	response to molecule of fungal origin	3
6.80E-08	GO:0008329	pattern recognition receptor activity	3
6.80E-08	GO:0009062	fatty acid catabolic process	3
6.80E-08	GO:0030101	natural killer cell activation	3
6.80E-08	GO:0045089	positive regulation of innate immune response	3
1.19E-07	GO:0034613	cellular protein localization	4
3.74E-07	GO:0001906	cell killing	4
4.58E-07	GO:0048268	clathrin coat assembly	3
1.02E-06	GO:0005925	focal adhesion	12
1.06E-06	GO:0015085	calcium ion transmembrane transporter activity	4
1.06E-06	GO:0017046	peptide hormone binding	4
1.36E-06	GO:0016055	Wnt receptor signaling pathway	11
1.64E-06	GO:0042542	response to hydrogen peroxide	7
1.77E-06	GO:0008203	cholesterol metabolic process	8
2.36E-06	GO:0005915	zonula adherens	3
2.36E-06	GO:0030122	AP-2 adaptor complex	3
2.36E-06	GO:0042832	defense response to protozoan	3
2.36E-06	GO:0051536	iron-sulfur cluster binding	3

2.41E-06	GO:0042632	cholesterol homeostasis	7
2.85E-06	GO:0005200	structural constituent of cytoskeleton	9
6.68E-06	GO:0042221	response to chemical stimulus	4
9.56E-06	GO:0003344	pericardium morphogenesis	2
9.56E-06	GO:0003401	axis elongation	2
9.56E-06	GO:0007620	copulation	2
9.56E-06	GO:0016362	activin receptor activity, type II	2
9.56E-06	GO:0042030	ATPase inhibitor activity	2
9.56E-06	GO:0045836	positive regulation of meiosis	2
9.56E-06	GO:0060029	convergent extension involved in organogenesis	2
9.56E-06	GO:0060638	mesenchymal-epithelial cell signaling	2
9.56E-06	GO:0060686	negative regulation of prostatic bud formation	2
9.56E-06	GO:0060750	epithelial cell proliferation involved in mammary gland duct elongation	2
9.56E-06	GO:0090009	primitive streak formation	2
9.76E-06	GO:0005899	insulin receptor complex	3
9.76E-06	GO:0016944	RNA polymerase II transcription elongation factor activity	3
9.76E-06	GO:0048739	cardiac muscle fiber development	3
9.76E-06	GO:0090004	positive regulation of establishment of protein localization in plasma membrane	3
1.28E-05	GO:0016337	cell-cell adhesion	10
1.40E-05	GO:0006874	cellular calcium ion homeostasis	7
1.51E-05	GO:0033198	response to ATP	4
1.63E-05	GO:0008021	synaptic vesicle	5
5.51E-05	GO:0045121	membrane raft	11
0.000101848	GO:0006910	phagocytosis, recognition	3
0.000101848	GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	3
0.000101848	GO:0050690	regulation of defense response to virus by virus	3
0.000101848	GO:0050727	regulation of inflammatory response	3
0.000103423	GO:0005861	troponin complex	2
0.000103423	GO:0006776	vitamin A metabolic process	2
0.000103423	GO:0008038	neuron recognition	2
0.000103423	GO:0010800	positive regulation of peptidyl-threonine phosphorylation	2
0.000103423	GO:0016226	iron-sulfur cluster assembly	2
0.000103423	GO:0017002	activin receptor activity	2
0.000103423	GO:0021761	limbic system development	2
0.000103423	GO:0034187	apolipoprotein E binding	2

0.000103423	GO:0042312	regulation of vasodilation	2
0.000103423	GO:0042713	sperm ejaculation	2
0.000103423	GO:0050999	regulation of nitric-oxide synthase activity	2
0.000103423	GO:0060065	uterus development	2
0.000103423	GO:0060071	Wnt receptor signaling pathway, planar cell polarity pathway	2
0.000103423	GO:0060326	cell chemotaxis	2
0.000103423	GO:0060744	mammary gland branching involved in thelarche	2
0.000103423	GO:0090103	cochlea morphogenesis	2
0.000115864	GO:0015297	antiporter activity	5
0.000123984	GO:0004896	cytokine receptor activity	4
0.000123984	GO:0030315	T-tubule	4
0.000123984	GO:0048041	focal adhesion assembly	4
0.000192338	GO:0003779	actin binding	21
0.000421355	GO:0005901	caveola	8
0.000651526	GO:0006939	smooth muscle contraction	3
0.000663174	GO:0001518	voltage-gated sodium channel complex	2
0.000663174	GO:0004740	pyruvate dehydrogenase (acetyl-transferring) kinase activity	2
0.000663174	GO:0005246	calcium channel regulator activity	2
0.000663174	GO:0006027	glycosaminoglycan catabolic process	2
0.000663174	GO:0010873	positive regulation of cholesterol esterification	2
0.000663174	GO:0019788	NEDD8 ligase activity	2
0.000663174	GO:0030510	regulation of BMP signaling pathway	2
0.000663174	GO:0045080	positive regulation of chemokine biosynthetic process	2
0.000663174	GO:0045116	protein neddylation	2
0.000663174	GO:0060068	vagina development	2
0.0006631	GO:0071277	cellular response to calcium ion	2

74			
0.001137248	GO:0046907	intracellular transport	4
0.001435728	GO:0032728	positive regulation of interferon-beta production	3
0.001435728	GO:0035098	ESC/E(Z) complex	3
0.002627184	GO:0001666	response to hypoxia	13
0.00285152	GO:0004364	glutathione transferase activity	4
0.00285152	GO:0050699	WW domain binding	4
0.002936752	GO:0006911	phagocytosis, engulfment	3
0.00294672	GO:0007494	midgut development	2
0.00294672	GO:0010820	positive regulation of T cell chemotaxis	2
0.00294672	GO:0018106	peptidyl-histidine phosphorylation	2
0.00294672	GO:0019907	cyclin-dependent protein kinase activating kinase holoenzyme complex	2
0.00294672	GO:0034374	low-density lipoprotein particle remodeling	2
0.00294672	GO:0044224	juxtaparanode region of axon	2
0.00294672	GO:0048806	genitalia development	2
0.00294672	GO:0051546	keratinocyte migration	2
0.00294672	GO:0060612	adipose tissue development	2
0.00294672	GO:0061036	positive regulation of cartilage development	2
0.003408272	GO:0031225	anchored to membrane	8
0.003752224	GO:0042470	melanosome	9
0.004240768	GO:0005516	calmodulin binding	11
0.004346496	GO:0021766	hippocampus development	4
0.005632816	GO:0008023	transcription elongation factor complex	3
0.007091504	GO:0004871	signal transducer activity	18
0.007891184	GO:0004867	serine-type endopeptidase inhibitor activity	7
0.009451792	GO:0030057	desmosome	4

0.009507008	GO:0009653	anatomical structure morphogenesis	8
0.010021984	GO:0005109	frizzled binding	2

Appendix 3. List of differentially expressed genes between NSE and AK

Symbol	Chr	logFC	adj.P.Val	Description
WIF1	12	-4.219315383	1.65E-21	WNT inhibitory factor 1
PHYHIP	8	-2.723224136	2.59E-19	phytanoyl-CoA 2-hydroxylase interacting protein
BTC	4	-2.566776274	5.29E-18	betacellulin
SP8	7	-3.684368433	4.30E-17	Sp8 transcription factor
RAB3B	1	-1.686256632	7.43E-15	RAB3B, member RAS oncogene family
TNNC1	3	-1.99616603	8.13E-15	troponin C type 1 (slow)
CCL27	9	-3.784777844	1.08E-14	chemokine (C-C motif) ligand 27
STMN2	8	2.903572234	1.60E-14	stathmin-like 2
LEPR	1	-2.27096095	3.40E-14	leptin receptor
SYT17	16	-2.259664347	1.69E-13	synaptotagmin XVII
NA	-	-1.906489491	2.42E-13	NA
CYP4B1	1	-3.000085113	3.36E-13	cytochrome P450, family 4, subfamily B, polypeptide 1
SGSM1	22	-0.971420925	1.09E-11	small G protein signaling modulator 1
PAMR1	11	-2.396125888	1.10E-11	peptidase domain containing associated with muscle regeneration 1
TPPP	5	-1.980729801	1.51E-11	tubulin polymerization promoting protein
DSC2	18	2.222951605	2.17E-11	desmocollin 2
SUSD2	22	-2.379402544	2.95E-11	sushi domain containing 2
PTPN21	14	-1.464634331	6.87E-11	protein tyrosine phosphatase, non-receptor type 21
NETO2	16	2.540280679	1.07E-10	neuropilin (NRP) and tolloid (TLL)-like 2
WDR66	12	1.890740818	1.51E-10	WD repeat domain 66
IRAK1	X	0.912036063	4.33E-10	interleukin-1 receptor-associated kinase 1
FBLIM1	1	1.600072208	4.33E-10	filamin binding LIM protein 1
CLDN23	8	-1.551793115	8.79E-10	claudin 23
KRT6B	12	2.139561146	1.01E-09	keratin 6B
IFI27	14	2.349345424	1.07E-09	interferon, alpha-inducible protein 27
S100A9	1	3.59527943	1.34E-09	S100 calcium binding protein A9
C7orf59	7	-0.980354962	1.40E-09	chromosome 7 open reading frame 59
GALNT6	12	2.227691829	1.47E-09	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)
SOX5	12	-1.982365486	2.97E-09	SRY (sex determining region Y)-box 5
CMAH	6	-1.661418146	3.17E-09	cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMP-N-acetylneuraminate monooxygenase) pseudogene
GATM	15	-1.318654326	3.73E-09	glycine amidinotransferase (L-arginine:glycine amidinotransferase)

CD24	6	1.612510574	4.79E-09	CD24 molecule
C4orf36	4	-1.821983315	4.92E-09	chromosome 4 open reading frame 36
MYO1B	2	1.291192726	4.94E-09	myosin IB
KRT16	17	3.253946738	5.82E-09	keratin 16
S100A2	1	0.836421013	6.03E-09	S100 calcium binding protein A2
ENAH	1	2.040161851	9.71E-09	enabled homolog (Drosophila)
SNTB1	8	-1.349195013	1.03E-08	syntrophin, beta 1 (dystrophin-associated protein A1, 59kDa, basic component 1)
AQP9	15	-1.845478701	1.19E-08	aquaporin 9
S100A8	1	2.342484273	1.24E-08	S100 calcium binding protein A8
AK7	14	-1.681348413	1.48E-08	adenylate kinase 7
DLG2	11	-1.933191684	1.76E-08	discs, large homolog 2 (Drosophila)
OASL	12	2.802527	2.29E-08	2'-5'-oligoadenylate synthetase-like
SLC8A1	2	-1.722647752	2.86E-08	solute carrier family 8 (sodium/calcium exchanger), member 1
BFSP1	20	1.470967319	3.23E-08	beaded filament structural protein 1, filensin
WNT16	7	-1.499361377	4.87E-08	wingless-type MMTV integration site family, member 16
GJB2	13	1.313995421	4.99E-08	gap junction protein, beta 2, 26kDa
PLLP	16	-1.776363663	6.11E-08	plasma membrane proteolipid (plasmolipin)
POU2F3	11	-1.396904523	6.90E-08	POU class 2 homeobox 3
GJB4	1	-1.451039476	7.01E-08	gap junction protein, beta 4, 30.3kDa
C1QTNF7	4	-1.604042815	7.08E-08	C1q and tumor necrosis factor related protein 7
WNT5A	3	2.319730967	7.77E-08	wingless-type MMTV integration site family, member 5A
ADH1B	4	-2.470779535	7.77E-08	alcohol dehydrogenase 1B (class I), beta polypeptide
C1orf135	1	1.514852991	1.01E-07	chromosome 1 open reading frame 135
TPBG	6	1.284329154	1.05E-07	trophoblast glycoprotein
COMP	19	2.447918217	1.11E-07	cartilage oligomeric matrix protein
C14orf132	14	-1.275570516	1.36E-07	chromosome 14 open reading frame 132
IGFL1	19	3.4012611	1.66E-07	IGF-like family member 1
KCNJ15	21	1.633086491	1.79E-07	potassium inwardly-rectifying channel, subfamily J, member 15
CLEC7A	12	2.030406535	2.15E-07	C-type lectin domain family 7, member A
OAS1	12	1.779846981	2.43E-07	2',5'-oligoadenylate synthetase 1, 40/46kDa
IL1F7	2	-2.206243299	2.48E-07	interleukin 1 family, member 7 (zeta)
USP54	10	-0.898164344	2.61E-07	ubiquitin specific peptidase 54
RSPH1	21	-1.211354255	3.36E-07	radial spoke head 1 homolog (Chlamydomonas)
GPC2	7	-1.092117918	3.36E-07	glypican 2
ALDH3A2	17	-1.203237996	3.39E-07	aldehyde dehydrogenase 3 family, member A2
RGMB	5	-1.023182889	3.69E-07	RGM domain family, member B
C9orf122	9	-1.651774126	3.73E-07	chromosome 9 open reading frame 122
GSTA3	6	-2.03753085	3.85E-07	glutathione S-transferase alpha 3
KLK1	19	-1.051024035	3.88E-07	kallikrein 1
TNFRSF21	6	1.9808069	3.94E-07	tumor necrosis factor receptor superfamily, member 21

GAL3ST4	7	-1.108617438	4.29E-07	galactose-3-O-sulfotransferase 4
MYOM2	8	-0.949774074	4.35E-07	myomesin (M-protein) 2, 165kDa
GPR177	1	-0.942933109	4.82E-07	G protein-coupled receptor 177
FAM70A	X	-1.184394342	5.22E-07	family with sequence similarity 70, member A
CYP4F12	19	-1.365839946	5.60E-07	cytochrome P450, family 4, subfamily F, polypeptide 12
PARD3	10	-0.952938896	5.76E-07	par-3 partitioning defective 3 homolog (C. elegans)
TPPP3	16	-1.519503982	5.92E-07	tubulin polymerization-promoting protein family member 3
KCTD17	22	-0.812674759	6.15E-07	potassium channel tetramerisation domain containing 17
KDM5A	12	-0.809783527	7.23E-07	lysine (K)-specific demethylase 5A
MAMDC2	9	-1.720624428	7.32E-07	MAM domain containing 2
SIPA1L2	1	1.232950993	7.32E-07	signal-induced proliferation-associated 1 like 2
CNGA1	4	-1.116728665	7.64E-07	cyclic nucleotide gated channel alpha 1
OAS2	12	2.172473846	8.27E-07	2'-5'-oligoadenylate synthetase 2, 69/71kDa
GATA3	10	-1.681693372	8.27E-07	GATA binding protein 3
SRGAP2P1	1	-1.072893665	8.27E-07	SLIT-ROBO Rho GTPase activating protein 2 pseudogene 1
TMEM99	17	-1.267083668	8.58E-07	transmembrane protein 99
C4orf31	4	-1.840488526	8.66E-07	chromosome 4 open reading frame 31
SLAMF7	1	2.203178473	8.73E-07	SLAM family member 7
NPY1R	4	-1.437421985	8.97E-07	neuropeptide Y receptor Y1
EEF2K	16	-1.073626963	9.64E-07	eukaryotic elongation factor-2 kinase
KRT9	17	3.866831475	1.08E-06	keratin 9
LYVE1	11	-1.694244949	1.09E-06	lymphatic vessel endothelial hyaluronan receptor 1
SAMD9	7	1.684078511	1.36E-06	sterile alpha motif domain containing 9
SLC1A2	11	-0.99863391	1.40E-06	solute carrier family 1 (glial high affinity glutamate transporter), member 2
SLC47A1	17	-1.419630899	1.55E-06	solute carrier family 47, member 1
MAP4K4	2	0.845137296	1.61E-06	mitogen-activated protein kinase kinase kinase kinase 4
RXRA	9	-0.776589712	1.61E-06	retinoid X receptor, alpha
DENND4C	9	-0.819556405	1.61E-06	DENN/MADD domain containing 4C
TMEM198	2	0.935810178	1.63E-06	transmembrane protein 198
TMEM132B	12	-0.817122463	1.65E-06	transmembrane protein 132B
KIAA0802	18	1.690244688	1.70E-06	KIAA0802
PALLD	4	0.812398774	1.96E-06	palladin, cytoskeletal associated protein
PITX1	5	1.476500875	1.96E-06	paired-like homeodomain 1
SCIN	7	-1.498386551	2.03E-06	scinderin
DIO2	14	1.410794715	2.28E-06	deiodinase, iodothyronine, type II
NP	14	1.261825123	2.54E-06	nucleoside phosphorylase
C3orf52	3	-1.446020377	2.58E-06	chromosome 3 open reading frame 52
IGF1	12	1.642505988	2.60E-06	insulin-like growth factor 1 (somatomedin C)

SYNE1	6	-0.923470207	2.60E-06	spectrin repeat containing, nuclear envelope 1
C11orf67	11	-0.71166701	2.64E-06	chromosome 11 open reading frame 67
SRXN1	20	1.226778481	2.64E-06	sulfiredoxin 1 homolog (S. cerevisiae)
KRT6A	12	2.457025505	2.78E-06	keratin 6A
ENO1	1	0.832101885	2.99E-06	enolase 1, (alpha)
SCARA5	8	-2.13153296	3.26E-06	scavenger receptor class A, member 5 (putative)
GTF3C6	6	0.545359787	3.26E-06	general transcription factor IIIC, polypeptide 6, alpha 35kDa
FAM126A	7	1.172415743	3.50E-06	family with sequence similarity 126, member A
DSG3	18	1.326400769	3.53E-06	desmoglein 3 (pemphigus vulgaris antigen)
ABCC1	16	0.953360832	3.53E-06	ATP-binding cassette, sub-family C (CFTR/MRP), member 1
CSRP2	12	1.565678105	3.58E-06	cysteine and glycine-rich protein 2
PRODH	22	-1.301358419	3.59E-06	proline dehydrogenase (oxidase) 1
TXNDC17	17	1.053771403	3.61E-06	thioredoxin domain containing 17
ADAM23	2	2.279633922	3.75E-06	ADAM metallopeptidase domain 23
C15orf59	15	-1.277842281	3.82E-06	chromosome 15 open reading frame 59
PRKCB	16	-1.109873417	3.91E-06	protein kinase C, beta
FOXN3	14	-0.664094	4.01E-06	forkhead box N3
FAM65C	20	1.627708704	4.08E-06	family with sequence similarity 65, member C
ZFYVE21	14	-0.648223034	4.44E-06	zinc finger, FYVE domain containing 21
N4BP2L1	13	-1.220474683	4.52E-06	NEDD4 binding protein 2-like 1
ADAMTSL3	15	-1.326651735	4.53E-06	ADAMTS-like 3
HUNK	21	-1.158573188	4.62E-06	hormonally up-regulated Neu-associated kinase
ZNF323	6	1.222242953	4.86E-06	zinc finger protein 323
SLC20A1	2	0.985405179	4.86E-06	solute carrier family 20 (phosphate transporter), member 1
FUT3	19	1.535498932	5.02E-06	fucosyltransferase 3 (galactoside 3(4)-L-fucosyltransferase, Lewis blood group)
C12orf5	12	1.233765685	5.07E-06	chromosome 12 open reading frame 5
ACVR2A	2	-0.926137449	5.42E-06	activin A receptor, type IIA
KIAA1467	12	-1.615435124	5.73E-06	KIAA1467
TXNDC5	6	0.79115768	5.80E-06	thioredoxin domain containing 5 (endoplasmic reticulum)
KRT17	17	1.689164585	5.88E-06	keratin 17
CMPK2	2	1.918473793	5.88E-06	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial
THY1	11	1.555223084	6.31E-06	Thy-1 cell surface antigen
MPZL1	1	0.787156482	6.71E-06	myelin protein zero-like 1
PRSS23	11	1.298776439	6.86E-06	protease, serine, 23
CNTNAP3	9	-1.694815043	6.90E-06	contactin associated protein-like 3
IKZF2	2	-1.602292506	6.90E-06	IKAROS family zinc finger 2 (Helios)
NDFIP2	13	-0.95975278	7.16E-06	Nedd4 family interacting protein 2
PRO0471	3	-1.149151688	7.16E-06	hypothetical LOC28994
EIF5	14	0.570378219	8.19E-06	eukaryotic translation initiation factor 5
YKT6	7	0.855689397	8.84E-06	YKT6 v-SNARE homolog (S. cerevisiae)

FAM49B	8	0.728013526	8.86E-06	family with sequence similarity 49, member B
HK2	2	1.05462202	9.17E-06	hexokinase 2
LOC643650	10	-1.420576304	9.41E-06	hypothetical protein LOC643650
FAM164A	8	-1.041564905	9.50E-06	family with sequence similarity 164, member A
ZSCAN18	19	-1.268761562	1.02E-05	zinc finger and SCAN domain containing 18
NPY5R	4	-0.717220478	1.03E-05	neuropeptide Y receptor Y5
LTA4H	12	-0.693645845	1.03E-05	leukotriene A4 hydrolase
RNF180	5	-1.438278345	1.08E-05	ring finger protein 180
SNCG	10	-0.796416312	1.11E-05	synuclein, gamma (breast cancer-specific protein 1)
PAIP2B	2	-1.307247416	1.11E-05	poly(A) binding protein interacting protein 2B
C20orf24	20	0.814091892	1.26E-05	chromosome 20 open reading frame 24
WFS1	4	-0.683505438	1.31E-05	Wolfram syndrome 1 (wolframin)
NEFL	8	3.563809174	1.31E-05	neurofilament, light polypeptide
SCD	10	1.473211644	1.35E-05	stearoyl-CoA desaturase (delta-9-desaturase)
IFI6	1	2.084578733	1.40E-05	interferon, alpha-inducible protein 6
ADAMDEC1	8	2.303583223	1.53E-05	ADAM-like, decysin 1
C1orf59	1	1.261547967	1.59E-05	chromosome 1 open reading frame 59
SRGAP2	1	-0.823368986	1.60E-05	SLIT-ROBO Rho GTPase activating protein 2
HIF1A	14	0.988575604	1.62E-05	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
ARL4C	2	0.861624053	1.62E-05	ADP-ribosylation factor-like 4C
SLC7A5	16	1.331999362	1.64E-05	solute carrier family 7 (cationic amino acid transporter, y+ system), member 5
DEFB4	8	4.24312331	1.69E-05	defensin, beta 4
ZDHHC11	5	-0.973551038	1.75E-05	zinc finger, DHHC-type containing 11
LRRC20	10	1.162441208	1.75E-05	leucine rich repeat containing 20
IFI44	1	1.874346514	1.75E-05	interferon-induced protein 44
AP2S1	19	0.989982492	1.75E-05	adaptor-related protein complex 2, sigma 1 subunit
ACOT8	20	-1.126086831	1.79E-05	acyl-CoA thioesterase 8
MYH14	19	-1.056965442	1.81E-05	myosin, heavy chain 14
ATOH8	2	-0.977092297	1.83E-05	atonal homolog 8 (Drosophila)
TMEM43	3	-0.696412265	1.96E-05	transmembrane protein 43
IL7	8	-1.292088499	1.96E-05	interleukin 7
ELL3	15	-0.84484672	1.96E-05	elongation factor RNA polymerase II-like 3
GLTSCR2	19	-0.649183801	1.96E-05	glioma tumor suppressor candidate region gene 2
SGK3	8	-0.947213565	2.04E-05	serum/glucocorticoid regulated kinase family, member 3
ZNF658	9	-0.736340566	2.05E-05	zinc finger protein 658
UBE2F	2	0.841113104	2.05E-05	ubiquitin-conjugating enzyme E2F (putative)
ZNF418	19	-1.128254365	2.05E-05	zinc finger protein 418
ID4	6	-1.362570731	2.18E-05	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
ZNF204	6	-1.253631153	2.18E-05	zinc finger protein 204 pseudogene

BCL2	18	-1.259152522	2.24E-05	B-cell CLL/lymphoma 2
LOC648987	5	0.897217074	2.31E-05	hypothetical LOC648987
AGFG2	7	-0.879381116	2.34E-05	ArfGAP with FG repeats 2
C7orf10	7	1.12474032	2.38E-05	chromosome 7 open reading frame 10
F2R	5	-1.017073706	2.46E-05	coagulation factor II (thrombin) receptor
SCEL	13	-2.174886732	2.47E-05	sciellin
GREM2	1	-0.957062849	2.50E-05	gremlin 2, cysteine knot superfamily, homolog (Xenopus laevis)
RHOT1	17	-0.738534469	2.54E-05	ras homolog gene family, member T1
ETV7	6	-1.383878495	2.56E-05	ets variant 7
MX1	21	1.732484456	2.59E-05	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)
INPP5A	10	-0.711644897	2.61E-05	inositol polyphosphate-5-phosphatase, 40kDa
EBNA1BP2	1	0.528910812	2.61E-05	EBNA1 binding protein 2
MID1	X	0.897925333	2.68E-05	midline 1 (Opitz/BBB syndrome)
EMP1	12	-1.113356273	2.70E-05	epithelial membrane protein 1
SYNCRIP	6	0.737268696	2.73E-05	synaptotagmin binding, cytoplasmic RNA interacting protein
C2orf67	2	-0.918301066	2.73E-05	chromosome 2 open reading frame 67
LMO7	13	-0.915816105	2.73E-05	LIM domain 7
PPP4R1	18	0.74107727	2.73E-05	protein phosphatase 4, regulatory subunit 1
EPHX2	8	-0.894923191	2.76E-05	epoxide hydrolase 2, cytoplasmic
CLDN11	3	-1.566806527	2.76E-05	claudin 11
GSTM5	1	-1.285096379	2.94E-05	glutathione S-transferase mu 5
MLLT11	1	1.40448242	3.01E-05	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 11
DZIP1L	3	-0.812694347	3.05E-05	DAZ interacting protein 1-like
GIPC2	1	-1.643298792	3.11E-05	GIPC PDZ domain containing family, member 2
VSNL1	2	1.084108398	3.16E-05	visinin-like 1
LIPA	10	0.715655706	3.25E-05	lipase A, lysosomal acid, cholesterol esterase
CBX2	17	1.063116422	3.31E-05	chromobox homolog 2 (Pc class homolog, Drosophila)
LASS6	2	-0.952986163	3.33E-05	LAG1 homolog, ceramide synthase 6
N4BP2L2	13	-0.56268256	3.33E-05	NEDD4 binding protein 2-like 2
PNRC1	6	-0.527615837	3.64E-05	proline-rich nuclear receptor coactivator 1
BZW1	2	0.57891889	3.66E-05	basic leucine zipper and W2 domains 1
ANKRD18A	9	-1.14711812	3.66E-05	ankyrin repeat domain 18A
CRIP1	14	-1.687982033	3.77E-05	cysteine-rich protein 1 (intestinal)
RAB31	18	1.062321588	3.82E-05	RAB31, member RAS oncogene family
PI3	20	2.790989125	3.88E-05	peptidase inhibitor 3, skin-derived
TNFRSF10B	8	0.934090139	3.94E-05	tumor necrosis factor receptor superfamily, member 10b
FMNL2	2	0.98327065	3.94E-05	formin-like 2
TMEM49	17	1.351731577	3.99E-05	transmembrane protein 49

RELL1	4	-0.78272333	4.00E-05	RELT-like 1
GJB6	13	1.48765847	4.19E-05	gap junction protein, beta 6, 30kDa
SPINK6	5	3.875208272	4.22E-05	serine peptidase inhibitor, Kazal type 6
LOC643827	9	-1.389112034	4.28E-05	similar to cell recognition molecule CASPR3
GCET2	3	-1.156197605	4.39E-05	germinal center expressed transcript 2
CTNND1	11	-0.749821965	4.46E-05	catenin (cadherin-associated protein), delta 1
MXD1	2	0.970792058	4.55E-05	MAX dimerization protein 1
CES1	16	-1.241229208	4.63E-05	carboxylesterase 1 (monocyte/macrophage serine esterase 1)
FGFR3	4	-0.980242284	4.64E-05	fibroblast growth factor receptor 3
RASAL2	1	-0.705599039	4.73E-05	RAS protein activator like 2
LMBRD1	6	-0.547656424	4.75E-05	LMBR1 domain containing 1
ACADL	2	-0.982992006	4.82E-05	acyl-Coenzyme A dehydrogenase, long chain
TMC8	17	0.994419548	4.83E-05	transmembrane channel-like 8
RPL13	16	-0.53256886	4.91E-05	ribosomal protein L13
ZNF677	19	-1.007706414	5.26E-05	zinc finger protein 677
ANG	14	-0.982791559	5.26E-05	angiogenin, ribonuclease, RNase A family, 5
CTPS	1	0.805391343	5.34E-05	CTP synthase
FOXE1	9	1.879910169	5.34E-05	forkhead box E1 (thyroid transcription factor 2)
GALNTL1	14	-1.48864083	5.36E-05	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 1
OBFC1	10	0.764686637	5.38E-05	oligonucleotide/oligosaccharide-binding fold containing 1
DENND1A	9	1.013513481	5.50E-05	DENN/MADD domain containing 1A
EPN2	17	-0.666432193	5.58E-05	epsin 2
ASAH1	8	-0.658586556	5.58E-05	N-acylsphingosine amidohydrolase (acid ceramidase) 1
SLC30A9	4	0.933845156	5.60E-05	solute carrier family 30 (zinc transporter), member 9
GTPBP4	10	0.645442007	5.61E-05	GTP binding protein 4
TMEM19	12	-0.812772988	5.64E-05	transmembrane protein 19
PGAP1	2	-0.807492606	5.71E-05	post-GPI attachment to proteins 1
BBS2	16	-0.748505371	5.71E-05	Bardet-Biedl syndrome 2
HOMER1	5	0.992829208	5.75E-05	homer homolog 1 (Drosophila)
FLJ32255	5	1.506250462	5.75E-05	hypothetical protein LOC643977
CTSC	11	0.829848209	5.75E-05	cathepsin C
STX7	6	-0.509832943	5.88E-05	syntaxin 7
DLX2	2	2.239148577	6.00E-05	distal-less homeobox 2
BCMO1	16	-1.047941839	6.11E-05	beta-carotene 15,15'-monooxygenase 1
CACNB2	10	-0.853287147	6.16E-05	calcium channel, voltage-dependent, beta 2 subunit
TMEM119	12	1.219422577	6.32E-05	transmembrane protein 119
PITX2	4	1.425008042	6.47E-05	paired-like homeodomain 2
TNFSF9	19	1.28187161	6.60E-05	tumor necrosis factor (ligand) superfamily, member 9

ZNF273	7	-1.050336147	6.93E-05	zinc finger protein 273
GDA	9	3.003024332	7.11E-05	guanine deaminase
ARHGAP23	17	-0.822234333	7.11E-05	Rho GTPase activating protein 23
GLCCI1	7	-1.330668512	7.22E-05	glucocorticoid induced transcript 1
PDK4	7	-1.126383593	7.39E-05	pyruvate dehydrogenase kinase, isozyme 4
HNRPDL	4	-0.743594774	7.58E-05	heterogeneous nuclear ribonucleoprotein D-like
CAB39L	13	-0.85196845	7.58E-05	calcium binding protein 39-like
SLC25A27	6	-0.841397394	7.79E-05	solute carrier family 25, member 27
HERC6	4	1.683644313	8.01E-05	hect domain and RLD 6
POLR3G	5	1.103392334	8.01E-05	polymerase (RNA) III (DNA directed) polypeptide G (32kD)
MAPRE1	20	0.537458304	8.05E-05	microtubule-associated protein, RP/EB family, member 1
CYP39A1	6	-1.447370086	8.17E-05	cytochrome P450, family 39, subfamily A, polypeptide 1
GNB4	3	-0.979380437	8.19E-05	guanine nucleotide binding protein (G protein), beta polypeptide 4
NDRG4	16	1.625795155	8.24E-05	NDRG family member 4
WDR1	4	0.634003715	8.25E-05	WD repeat domain 1
TNFRSF19	13	-1.1034056	8.39E-05	tumor necrosis factor receptor superfamily, member 19
KRT77	12	-1.939033838	8.76E-05	keratin 77
TMEM64	8	0.752127494	8.76E-05	transmembrane protein 64
CNTNAP2	7	2.283007504	8.85E-05	contactin associated protein-like 2
KIF21A	12	-0.913919767	8.97E-05	kinesin family member 21A
CD55	1	-1.016394607	9.00E-05	CD55 molecule, decay accelerating factor for complement (Cromer blood group)
MKRN1	7	-0.62970988	9.10E-05	makorin ring finger protein 1
CLDND1	3	-0.768114494	9.10E-05	claudin domain containing 1
PITPNC1	17	1.108052369	9.24E-05	phosphatidylinositol transfer protein, cytoplasmic 1
NELL2	12	2.354454276	9.30E-05	NEL-like 2 (chicken)
SPRY4	5	1.297283639	9.34E-05	sprouty homolog 4 (Drosophila)
FRMD4A	10	-0.775832194	9.34E-05	FERM domain containing 4A
RFTN1	3	1.401497423	9.77E-05	raftlin, lipid raft linker 1

Appendix 4. Full list of significantly overrepresented GO terms comparing AK and NSE.

Adj. pvalue	**GO**	**Name**	**Probes**
<0.0001	GO:0001959	regulation of cytokine-mediated signaling pathway	7
<0.0001	GO:0032597	B cell receptor transport into membrane raft	6
<0.0001	GO:0032600	chemokine receptor transport out of membrane raft	6
<0.0001	GO:0032913	negative regulation of transforming growth factor-beta3 production	6
<0.0001	GO:0002768	immune response-regulating cell surface receptor signaling pathway	6
<0.0001	GO:0001775	cell activation	6
<0.0001	GO:0043627	response to estrogen stimulus	15
<0.0001	GO:0004768	stearoyl-CoA 9-desaturase activity	5
<0.0001	GO:0003408	optic cup formation involved in camera-type eye development	4
<0.0001	GO:0032273	positive regulation of protein polymerization	4
<0.0001	GO:0030856	regulation of epithelial cell differentiation	6
<0.0001	GO:0043408	regulation of MAPKKK cascade	6
<0.0001	GO:0045730	respiratory burst	7
<0.0001	GO:0033210	leptin-mediated signaling pathway	4
<0.0001	GO:0002237	response to molecule of bacterial origin	6
<0.0001	GO:0031295	T cell costimulation	6
<0.0001	GO:0042104	positive regulation of activated T cell proliferation	6
<0.0001	GO:0001874	zymosan receptor activity	4
<0.0001	GO:0001879	detection of yeast	4
<0.0001	GO:0031669	cellular response to nutrient levels	4
<0.0001	GO:0051346	negative regulation of hydrolase activity	4
<0.0001	GO:0030296	protein tyrosine kinase activator activity	6
<0.0001	GO:0042325	regulation of phosphorylation	6
<0.0001	GO:0016337	cell-cell adhesion	16
<0.0001	GO:0048873	homeostasis of number of cells within a tissue	5
<0.0001	GO:0001730	2'-5'-oligoadenylate synthetase activity	3
<0.0001	GO:0031115	negative regulation of microtubule polymerization	3
<0.0001	GO:0001846	opsonin binding	4
<0.0001	GO:0002752	cell surface pattern recognition receptor signaling pathway	4
<0.0001	GO:0009756	carbohydrate mediated signaling	4
<0.0001	GO:0031334	positive regulation of protein complex assembly	4
<0.0001	GO:0042287	MHC protein binding	4
<0.0001	GO:0046689	response to mercury ion	4
<0.0001	GO:0035094	response to nicotine	6

<0.0001	GO:0051924	regulation of calcium ion transport	6
<0.0001	GO:0005432	calcium:sodium antiporter activity	4
<0.0001	GO:0060401	cytosolic calcium ion transport	4
<0.0001	GO:0015068	glycine amidinotransferase activity	3
<0.0001	GO:0051371	muscle alpha-actinin binding	3
<0.0001	GO:0051902	negative regulation of mitochondrial depolarization	3
<0.0001	GO:0002238	response to molecule of fungal origin	4
<0.0001	GO:0006590	thyroid hormone generation	4
<0.0001	GO:0008329	pattern recognition receptor activity	4
<0.0001	GO:0016500	protein-hormone receptor activity	4
<0.0001	GO:0046785	microtubule polymerization	4
<0.0001	GO:0006601	creatine biosynthetic process	3
<0.0001	GO:0060509	Type I pneumocyte differentiation	3
<0.0001	GO:0001542	ovulation from ovarian follicle	4
<0.0001	GO:0002028	regulation of sodium ion transport	4
<0.0001	GO:0015081	sodium ion transmembrane transporter activity	4
<0.0001	GO:0007204	elevation of cytosolic calcium ion concentration	11
<0.0001	GO:0042755	eating behavior	5
<0.0001	GO:0042832	defense response to protozoan	4
<0.0001	GO:0001601	peptide YY receptor activity	2
<0.0001	GO:0001602	pancreatic polypeptide receptor activity	2
<0.0001	GO:0002741	positive regulation of cytokine secretion involved in immune response	2
<0.0001	GO:0003103	positive regulation of diuresis	2
<0.0001	GO:0003323	pancreatic B cell development	2
<0.0001	GO:0004396	hexokinase activity	2
<0.0001	GO:0004686	elongation factor-2 kinase activity	2
<0.0001	GO:0004983	neuropeptide Y receptor activity	2
<0.0001	GO:0005110	frizzled-2 binding	2
<0.0001	GO:0005115	receptor tyrosine kinase-like orphan receptor binding	2
<0.0001	GO:0030825	positive regulation of cGMP metabolic process	2
<0.0001	GO:0034755	iron ion transmembrane transport	2
<0.0001	GO:0048186	inhibin beta-A binding	2
<0.0001	GO:0048187	inhibin beta-B binding	2
<0.0001	GO:0048850	hypophysis morphogenesis	2
<0.0001	GO:0060011	Sertoli cell proliferation	2
<0.0001	GO:0060067	cervix development	2
<0.0001	GO:0070245	positive regulation of thymocyte apoptosis	2
<0.0001	GO:0070287	ferritin receptor activity	2
<0.0001	GO:0071542	dopaminergic neuron differentiation	2
<0.0001	GO:0004800	thyroxine 5'-deiodinase activity	3
<0.0001	GO:0018125	peptidyl-cysteine methylation	3
<0.0001	GO:0042404	thyroid hormone catabolic process	3
<0.0001	GO:0016055	Wnt receptor signaling pathway	13
<0.0001	GO:0001578	microtubule bundle formation	5
<0.0001	GO:0043406	positive regulation of MAP kinase activity	6
<0.0001	GO:0030175	filopodium	8
<0.0001	GO:0008037	cell recognition	5
<0.0001	GO:0006112	energy reserve metabolic process	4
<0.0001	GO:0006910	phagocytosis	4
<0.0001	GO:0002248	connective tissue replacement involved in inflammatory	2

		response wound healing	
<0.0001	GO:0004301	epoxide hydrolase activity	2
<0.0001	GO:0006582	melanin metabolic process	2
<0.0001	GO:0006808	regulation of nitrogen utilization	2
<0.0001	GO:0010559	regulation of glycoprotein biosynthetic process	2
<0.0001	GO:0014042	positive regulation of neuron maturation	2
<0.0001	GO:0021747	cochlear nucleus development	2
<0.0001	GO:0030338	CMP-N-acetylneuraminate monooxygenase activity	2
<0.0001	GO:0031117	positive regulation of microtubule depolymerization	2
<0.0001	GO:0032814	regulation of natural killer cell activation	2
<0.0001	GO:0032848	negative regulation of cellular pH reduction	2
<0.0001	GO:0043375	CD8-positive	2
<0.0001	GO:0046671	negative regulation of retinal cell programmed cell death	2
<0.0001	GO:0048087	positive regulation of developmental pigmentation	2
<0.0001	GO:0048743	positive regulation of skeletal muscle fiber development	2
<0.0001	GO:0048753	pigment granule organization	2
<0.0001	GO:0060164	regulation of timing of neuron differentiation	2
<0.0001	GO:0060760	positive regulation of response to cytokine stimulus	2
<0.0001	GO:0060762	regulation of branching involved in mammary gland duct morphogenesis	2
<0.0001	GO:0070059	apoptosis in response to endoplasmic reticulum stress	2
<0.0001	GO:0071346	cellular response to interferon-gamma	2
<0.0001	GO:0072201	negative regulation of mesenchymal cell proliferation	2
<0.0001	GO:0090037	positive regulation of protein kinase C signaling cascade	2
<0.0001	GO:0004028	3-chloroallyl aldehyde dehydrogenase activity	3
<0.0001	GO:0004030	aldehyde dehydrogenase [NAD(P)+] activity	3
<0.0001	GO:0007494	midgut development	3
<0.0001	GO:0008021	synaptic vesicle	6
<0.0001	GO:0008544	epidermis development	11
<0.0001	GO:0015631	tubulin binding	5
<0.0001	GO:0000085	G2 phase of mitotic cell cycle	4
<0.0001	GO:0008629	induction of apoptosis by intracellular signals	6
<0.0001	GO:0001514	selenocysteine incorporation	3
<0.0001	GO:0002360	T cell lineage commitment	3
<0.0001	GO:0030104	water homeostasis	3
<0.0001	GO:0030166	proteoglycan biosynthetic process	3
<0.0001	GO:0030054	cell junction	24
0.000	GO:0005886	plasma membrane	86
0.000	GO:0006911	phagocytosis	4
0.000	GO:0000015	phosphopyruvate hydratase complex	2
0.000	GO:0001667	ameboidal cell migration	2
0.000	GO:0002320	lymphoid progenitor cell differentiation	2
0.000	GO:0004634	phosphopyruvate hydratase activity	2
0.000	GO:0007442	hindgut morphogenesis	2
0.000	GO:0009441	glycolate metabolic process	2
0.000	GO:0010044	response to aluminum ion	2
0.000	GO:0014031	mesenchymal cell development	2
0.000	GO:0014834	satellite cell maintenance involved in skeletal muscle regeneration	2
0.000	GO:0014896	muscle hypertrophy	2

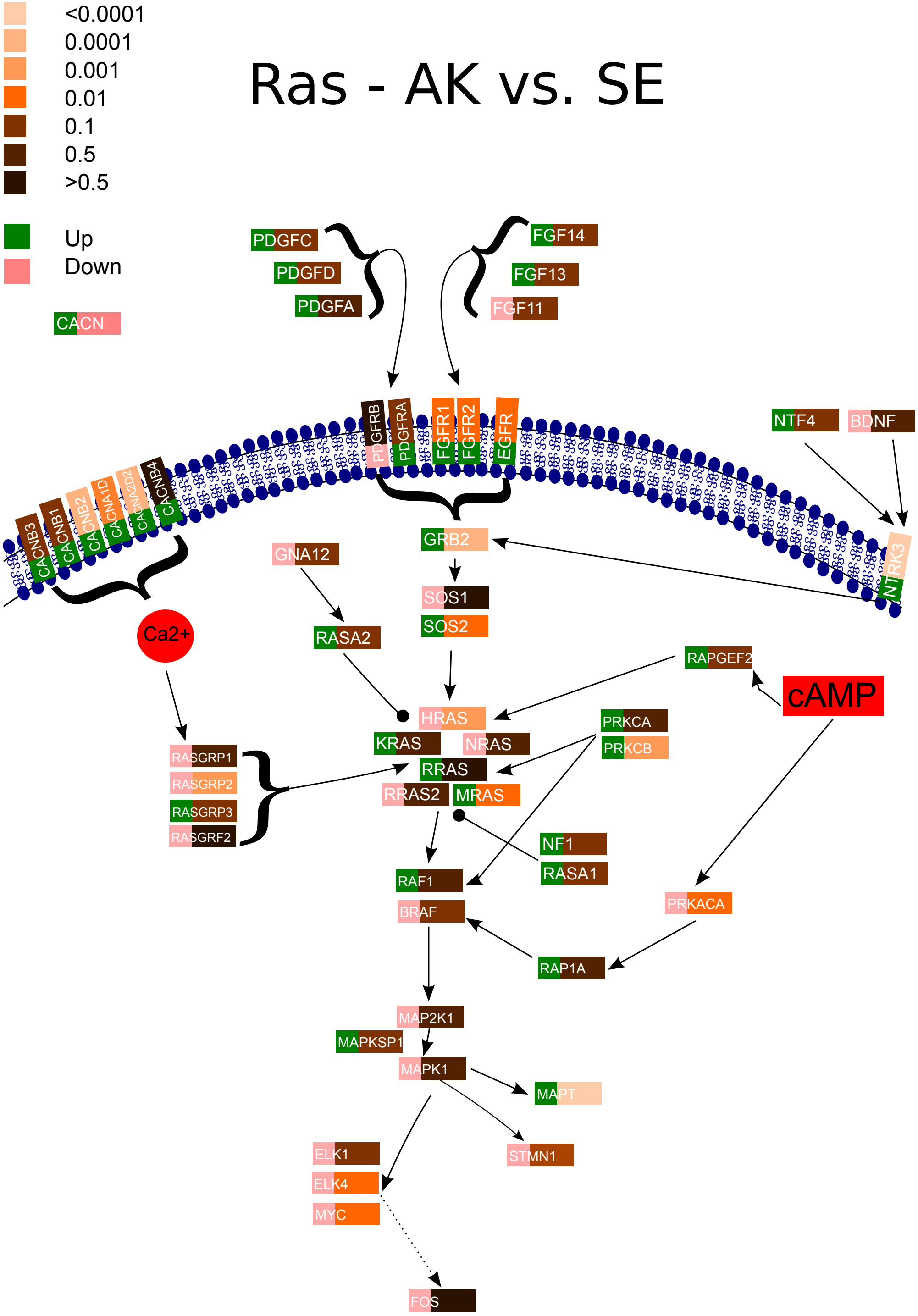
0.000	GO:0014904	myotube cell development	2
0.000	GO:0022612	gland morphogenesis	2
0.000	GO:0031014	troponin T binding	2
0.000	GO:0034673	inhibin-betaglycan-ActRII complex	2
0.000	GO:0043084	penile erection	2
0.000	GO:0045069	regulation of viral genome replication	2
0.000	GO:0046549	retinal cone cell development	2
0.000	GO:0060463	lung lobe morphogenesis	2
0.000	GO:0060599	lateral sprouting involved in mammary gland duct morphogenesis	2
0.000	GO:0060741	prostate gland stromal morphogenesis	2
0.000	GO:0048268	clathrin coat assembly	3
0.000	GO:0007584	response to nutrient	10
0.000	GO:0014911	positive regulation of smooth muscle cell migration	4
0.000	GO:0015085	calcium ion transmembrane transporter activity	4
0.000	GO:0017046	peptide hormone binding	4
0.000	GO:0045121	membrane raft	12
0.000	GO:0005149	interleukin-1 receptor binding	3
0.000	GO:0006600	creatine metabolic process	3
0.000	GO:0017157	regulation of exocytosis	3
0.000	GO:0030122	AP-2 adaptor complex	3
0.000	GO:0032469	endoplasmic reticulum calcium ion homeostasis	3
0.001	GO:0007223	Wnt receptor signaling pathway	4
0.001	GO:0030100	regulation of endocytosis	4
0.001	GO:0050699	WW domain binding	5
0.001	GO:0042221	response to chemical stimulus	4
0.001	GO:0003014	renal system process	2
0.001	GO:0003344	pericardium morphogenesis	2
0.001	GO:0003401	axis elongation	2
0.001	GO:0004024	alcohol dehydrogenase activity	2
0.001	GO:0007620	copulation	2
0.001	GO:0016362	activin receptor activity	2
0.001	GO:0030534	adult behavior	2
0.001	GO:0033033	negative regulation of myeloid cell apoptosis	2
0.001	GO:0045836	positive regulation of meiosis	2
0.001	GO:0046835	carbohydrate phosphorylation	2
0.001	GO:0051434	BH3 domain binding	2
0.001	GO:0060029	convergent extension involved in organogenesis	2
0.001	GO:0060638	mesenchymal-epithelial cell signaling	2
0.001	GO:0060686	negative regulation of prostatic bud formation	2
0.001	GO:0060750	epithelial cell proliferation involved in mammary gland duct elongation	2
0.001	GO:0090009	primitive streak formation	2
0.001	GO:0016338	calcium-independent cell-cell adhesion	3
0.001	GO:0005516	calmodulin binding	13
0.001	GO:0006139	nucleobase	9
0.001	GO:0008135	translation factor activity	4
0.001	GO:0032760	positive regulation of tumor necrosis factor production	4
0.001	GO:0033198	response to ATP	4
0.002	GO:0014068	positive regulation of phosphoinositide 3-kinase cascade	5

0.002	GO:0030057	desmosome	5
0.003	GO:0048146	positive regulation of fibroblast proliferation	6
0.003	GO:0048706	embryonic skeletal system development	5
0.003	GO:0045445	myoblast differentiation	4
0.003	GO:0004029	aldehyde dehydrogenase (NAD) activity	3
0.003	GO:0045821	positive regulation of glycolysis	3
0.003	GO:0001733	galactosylceramide sulfotransferase activity	1
0.003	GO:0003834	beta-carotene 15	1
0.003	GO:0004127	cytidylate kinase activity	1
0.004	GO:0004463	leukotriene-A4 hydrolase activity	1
0.005	GO:0004657	proline dehydrogenase activity	1
0.005	GO:0004731	purine-nucleoside phosphorylase activity	1
0.005	GO:0004771	sterol esterase activity	1
0.005	GO:0005139	interleukin-7 receptor binding	1
0.005	GO:0005275	amine transmembrane transporter activity	1
0.005	GO:0005345	purine transmembrane transporter activity	1
0.005	GO:0005350	pyrimidine transmembrane transporter activity	1
0.005	GO:0006148	inosine catabolic process	1
0.005	GO:0006863	purine transport	1
0.005	GO:0007529	establishment of synaptic specificity at neuromuscular junction	1
0.005	GO:0014707	branchiomeric skeletal muscle development	1
0.005	GO:0015166	polyol transmembrane transporter activity	1
0.005	GO:0015722	canalicular bile acid transport	1
0.005	GO:0015791	polyol transport	1
0.007	GO:0015855	pyrimidine transport	1
0.008	GO:0016401	palmitoyl-CoA oxidase activity	1

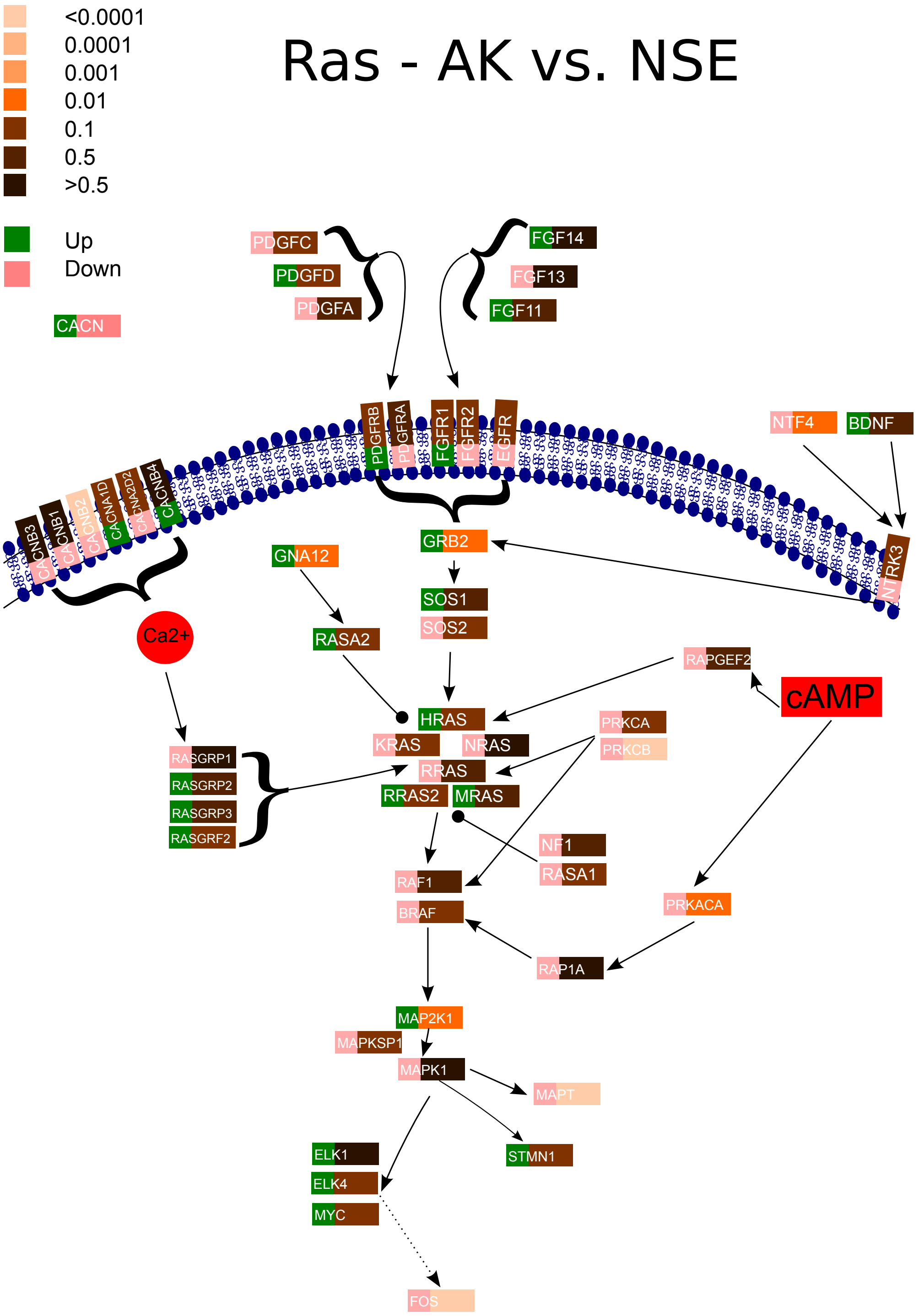
Appendix 5. The average expression level of EMT-associated markers in skin, AK and cSCC subtypes.

Symbol	NSE	SE	AK	WD	MD	MPD/PD
CDH26	3.170518799	3.177116975	4.059688057	4.231849863	3.448734376	3.833322814
CDH6	3.592186487	3.897070071	3.675378121	3.532581357	3.81366408	3.46229941
CDH19	4.403567759	5.0485751	4.150877039	3.262219893	3.071648697	3.0981883
CDH24	4.492057395	4.3760276	4.820957304	5.134630992	5.040684334	5.180777477
CLDN7	4.797037209	5.165910872	4.859272821	5.323079378	5.657939917	5.247951704
CDH11	5.204319666	5.716820891	5.96244819	5.188902018	5.184095444	6.757921243
ZEB2	5.613023876	5.940128129	5.767595503	5.043695386	5.493680956	5.775746809
MUC1	5.998372643	6.135153036	6.020093853	5.514333997	6.206169468	6.316493795
ZEB1	6.10271899	6.633469006	5.92906964	5.16438705	5.240731473	5.442788328
THBS1	6.316411937	6.73832734	6.910347733	6.178691523	7.136763558	7.37723831
ACTA2	6.450238905	6.882459304	6.302710882	5.10031612	5.179194309	5.059989259
CLDN5	6.526357012	6.326198891	6.191996379	5.715681189	5.581970554	5.711427368
CDH5	6.599603651	6.76305059	6.757345453	6.185365825	6.410189006	6.143613476
TWIST2	6.615820115	6.910274758	6.607720838	6.030882716	5.780019879	5.691582404
CLDN8	6.737020125	6.859036068	5.149662542	3.135776213	3.184871737	2.829829192
THBS3	6.975516949	7.063314246	6.659644943	6.308370851	6.498851921	6.341135178
CDH22	7.000105888	6.986488715	6.657042533	6.364158638	6.226154748	5.981568993
CLDN11	7.062131096	7.114870082	5.332968623	3.425846783	3.238459836	3.474215484
KRT18	7.079831145	7.512552806	6.988925506	7.389839345	8.423051339	7.868582444
VIM	7.155966312	7.340963189	7.047479392	6.387494029	6.661332718	7.113356251
CLDN12	7.173310834	7.157702405	7.41074766	8.120991374	8.207411782	8.060856293
KRT8	7.173794177	7.51373597	7.314437218	7.039684421	7.431681846	7.353694574
ACTG2	7.328634148	8.402112958	6.20259271	4.739848383	4.878860731	4.738136182
TCF3	7.634572469	6.98052364	7.848196341	7.379471938	7.540295199	7.799644015
TWIST1	7.778881863	8.112325734	8.092087352	6.950974208	6.8914464	7.881007571
CDH3	7.996803555	7.954140372	9.024104065	11.08459005	10.95878934	9.966808468
KRT9	8.052802944	7.557635933	11.76086034	7.212578966	6.91570135	8.865835797
FN1	8.181694549	8.542290938	7.938035099	7.828426345	8.231692184	8.98165685
FGFR2	8.279369451	8.311614765	7.943676128	7.50138596	7.060636872	5.81771095
CTNNB1	8.346254808	8.559817594	8.193543599	7.86784897	8.087911518	7.991561633
CDH1	9.3096953	9.167568361	9.275898308	9.83533432	9.324487461	8.37509494
SNAI2	9.332077914	9.308725643	9.640616902	10.09308508	10.15715279	10.04635047
OCLN	9.342358153	9.174887062	8.816605023	8.014716469	7.291643216	7.395167586
CTNND1	9.439118702	9.236009159	8.880216726	8.929163869	9.125666087	8.698512498
THBS2	10.45892661	10.73442033	10.99829583	10.79974622	9.925283459	10.00357589
CLDN1	11.58739888	11.16214191	10.50688777	10.25227875	10.08656634	9.579953086

Appendix 6. RAS/MAPK signaling in progression from SE to AK and NSE to AK



Ras - AK vs. NSE


<0.0001
0.0001
0.001
0.01
0.1
0.5
>0.5
Up
Down
CACN
PDGFC
PDGFD
PDGFA
FGF14
FGF13
FGF11
NTF4
BDNF
CACNB3
CACNB1
CACNB2
CACNA1D
CACNA2D2
CACNB4
PDGFRB
PDGFRA
FGFR1
FGFR2
EGFR
NTRK3
Ca2+
GNA12
GRB2
SOS1
SOS2
RASA2
RAPGEF2
cAMP
HRAS
KRAS
NRAS
RRAS
RRAS2
MRAS
PRKCA
PRKCB
RASGRP1
RASGRP2
RASGRP3
RASGRF2
NF1
RASA1
RAF1
BRAF
PRKACA
RAP1A
MAP2K1
MAPKSP1
MAPK1
MAPT
STMN1
ELK1
ELK4
MYC
FOS

Appendix 7. Differentially methylated genes between primary human keratinocytes and cultured primary cSCC cell lines.

SYMBOL	Description	Chr	adj.p("BH")
TRIM2	tripartite motif-containing 2	4	1.22531E-07
SPARCL1	SPARC-like 1 (mast9, hevin)	4	5.50704E-07
C1QTNF6	C1q and tumor necrosis factor related protein 6	22	5.50704E-07
AQP12A	aquaporin 12A	-	5.50704E-07
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	9	6.95756E-07
YPEL4	yippee-like 4 (Drosophila)	11	5.13957E-06
ECE1	endothelin converting enzyme 1	1	5.13957E-06
RGS17	regulator of G-protein signalling 17	6	6.69546E-06
FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)	8	8.67082E-06
BARHL2	BarH-like 2 (Drosophila)	1	2.32999E-05
CASP8	caspase 8, apoptosis-related cysteine peptidase	2	3.18007E-05
TM4SF19	transmembrane 4 L six family member 19	3	3.36019E-05
GABRA5	gamma-aminobutyric acid (GABA) A receptor, alpha 5	15	3.36019E-05
TRIM15	tripartite motif-containing 15	6	3.84449E-05
NFAM1	NFAT activating protein with ITAM motif 1	22	3.84449E-05
NEFH	neurofilament, heavy polypeptide 200kDa	22	4.32511E-05
SEC31L2	-	10	4.32511E-05
SH3TC1	SH3 domain and tetratricopeptide repeats 1	4	4.80662E-05
GAL3ST3	galactose-3-O-sulfotransferase 3	11	4.80662E-05
GDPD5	glycerophosphodiester phosphodiesterase domain containing 5	11	5.51624E-05
C18orf20	chromosome 18 open reading frame 20	18	5.70669E-05
MGC39545	-	11	5.70669E-05
TMEM22	transmembrane protein 22	3	6.36817E-05
GARNL3	GTPase activating Rap/RanGAP domain-like 3	9	7.0854E-05
PRKDC	protein kinase, DNA-activated, catalytic polypeptide	8	7.33318E-05
FXYD4	FXYD domain containing ion transport regulator 4	10	9.27401E-05
MBNL1	muscleblind-like (Drosophila)	3	9.71679E-05
ZDHHC11	zinc finger, DHHC-type containing 11	5	9.74256E-05
DDX49	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49	19	0.000112091
PPP3R2	protein phosphatase 3 (formerly 2B), regulatory subunit B, 19kDa, beta isoform (calcineurin B, type II)	9	0.000112091
MYH4	myosin, heavy chain 4, skeletal muscle	17	0.000112596
CR1	complement component (3b/4b) receptor 1 (Knops blood group)	1	0.000119724
NDRG2	NDRG family member 2	14	0.000136334

CNTNAP4	contactin associated protein-like 4	16	0.000136334
FAM50B	family with sequence similarity 50, member B	6	0.000136334
KIAA0427	KIAA0427	18	0.000146288
NIPSNAP1	nipsnap homolog 1 (C. elegans)	22	0.000150924
MPST	mercaptopyruvate sulfurtransferase	22	0.000150924
KYNU	kynureninase (L-kynurenine hydrolase)	2	0.000178931
QTRT1	queuine tRNA-ribosyltransferase 1 (tRNA-guanine transglycosylase)	19	0.00020842
TGIF2	TGFB-induced factor 2 (TALE family homeobox)	20	0.00026328
CMTM2	CKLF-like MARVEL transmembrane domain containing 2	16	0.000267798
STATH	statherin	4	0.000267798
GPR75	G protein-coupled receptor 75	2	0.000275936
FN3K	fructosamine 3 kinase	17	0.000277231
PRAME	preferentially expressed antigen in melanoma	22	0.000284809
TP53I3	tumor protein p53 inducible protein 3	2	0.000315786
C14orf166B	chromosome 14 open reading frame 166B	14	0.00032672
WFIKKN2	WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2	17	0.00035364
IGSF2	immunoglobulin superfamily, member 2	1	0.000373536
TM4SF11	-	16	0.000377574
C19orf19	chromosome 19 open reading frame 19	19	0.000433263
KRT7	keratin 7	12	0.000433263
EML1	echinoderm microtubule associated protein like 1	14	0.000433263
HSPCAL3	-	11	0.000433263
SERPINA3	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	14	0.000437043
NRTN	neurturin	19	0.000437043
B4GALT1	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1	9	0.000437043
GPR137	G protein-coupled receptor 137	11	0.000437043
LSAMP	limbic system-associated membrane protein	3	0.000463246
NRIP2	nuclear receptor interacting protein 2	12	0.000540445
UCHL1	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	4	0.000559268
ZIM2	zinc finger, imprinted 2	19	0.000573301
SAA2	serum amyloid A2	11	0.000573301
CHRNB4	cholinergic receptor, nicotinic, beta 4	15	0.00060761
C20orf179	chromosome 20 open reading frame 179	20	0.00060761
ACOT8	acyl-CoA thioesterase 8	-	0.000616779
KLHL21	kelch-like 21 (Drosophila)	1	0.000648082
TBC1D3C	TBC1 domain family, member 3C	17	0.000648082
TBX20	T-box 20	7	0.000669432
CHAT	choline acetyltransferase	10	0.000698374
THSD1	thrombospondin, type I, domain containing 1	13	0.000711337
GAGE4	G antigen 4	X	0.000751007
CHRNE	cholinergic receptor, nicotinic, epsilon	17	0.000756053
TMEM35	transmembrane protein 35	X	0.000761818
C21orf25	chromosome 21 open reading frame 25	21	0.000849069

PLCL1	phospholipase C-like 1	2	0.000860669
UNQ1940	-	7	0.000860669
SRRM2	serine/arginine repetitive matrix 2	16	0.000893946
MMP20	matrix metallopeptidase 20 (enamelysin)	11	0.000917033
TNFRSF11B	tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)	8	0.000917033
C9orf78	chromosome 9 open reading frame 78	9	0.000917033
ITK	IL2-inducible T-cell kinase	5	0.000967419
WFDC10B	WAP four-disulfide core domain 10B	20	0.001032763
PRKCB1	protein kinase C, beta 1	16	0.001078886
CCT6A	chaperonin containing TCP1, subunit 6A (zeta 1)	7	0.001100518
KLF3	Kruppel-like factor 3 (basic)	4	0.001114218
MAG	myelin associated glycoprotein	19	0.001169886
SIM1	single-minded homolog 1 (Drosophila)	6	0.001238979
PARC	-	6	0.001347767
ATP6V0A1	ATPase, H+ transporting, lysosomal V0 subunit a1	17	0.001367682
SST	somatostatin	3	0.001370235
ZIC1	Zic family member 1 (odd-paired homolog, Drosophila)	3	0.001371336
HBII-438B	-	15	0.001381504
KRTHB1	-	12	0.001392604
FLJ22555	-	2	0.001428385
FLJ39501	-	19	0.001428385
FMNL1	formin-like 1	17	0.001428385
CNN1	calponin 1, basic, smooth muscle	19	0.001465384
TNFRSF10C	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain	8	0.001465384
SYNGR1	synaptogyrin 1	22	0.001465384
LOC339789	-	2	0.001465384
F2RL3	coagulation factor II (thrombin) receptor-like 3	19	0.001520969
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	12	0.001605272
EVC2	Ellis van Creveld syndrome 2 (limbin)	4	0.001627908
TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B	1	0.001722942
FAM79B	family with sequence similarity 79, member B	3	0.001722942
FAM50B	family with sequence similarity 50, member B	6	0.001750441
RIMS3	regulating synaptic membrane exocytosis 3	1	0.001750441
TBX20	T-box 20	7	0.001750441
PPGB	protective protein for beta-galactosidase (galactosialidosis)	20	0.001750441
PCSK1	proprotein convertase subtilisin/kexin type 1	5	0.001750441
ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1	10	0.001750441
NA	neurocanthocytosis	-	0.001750441
SMPD3	sphingomyelin phosphodiesterase 3, neutral membrane (neutral sphingomyelinase II)	16	0.001750441
GCNT4	glucosaminyl (N-acetyl) transferase 4, core 2 (beta-1,6-N-acetylglucosaminyltransferase)	5	0.001793302
TBC1D3C	TBC1 domain family, member 3C	17	0.00180742
PCDHGB7	protocadherin gamma subfamily B, 7	5	0.00180742

FLJ46481	-	4	0.00180742
OR7C2	olfactory receptor, family 7, subfamily C, member 2	19	0.001841952
SKIP	-	2	0.001875265
DMRTC1	DMRT-like family C1	X	0.002067509
KRT25A	-	17	0.002124399
SNX9	sorting nexin 9	6	0.002145745
C10orf35	chromosome 10 open reading frame 35	10	0.002146706
FLJ00060	-	19	0.0021851
C22orf15	chromosome 22 open reading frame 15	22	0.002235644
MGC35295	-	11	0.002235644
SLC43A3	solute carrier family 43, member 3	11	0.002240191
RSHL1	radial spokehead-like 1	19	0.002281932
GATA4	GATA binding protein 4	8	0.00228307
PRLHR	prolactin releasing hormone receptor	10	0.002296916
LOC122258	-	13	0.002308457
TATDN1	TatD DNase domain containing 1	8	0.002323653
TMEM35	transmembrane protein 35	X	0.002339402
ZNF662	zinc finger protein 662	3	0.002363037
SPINK2	serine peptidase inhibitor, Kazal type 2 (acrosin-trypsin inhibitor)	4	0.002365943
TGM6	transglutaminase 6	20	0.002365943
CLEC4D	C-type lectin domain family 4, member D	12	0.002365943
FOXD2	forkhead box D2	1	0.002365943
ATP10A	ATPase, Class V, type 10A	15	0.002365943
SLC12A5	solute carrier family 12, (potassium-chloride transporter) member 5	20	0.002365943
REGL	regenerating islet-derived-like, pancreatic stone protein-like, pancreatic thread protein-like (rat)	2	0.002365943
ATP10A	ATPase, Class V, type 10A	15	0.002365943
AQP9	aquaporin 9	-	0.002382694
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	9	0.002396521
GGT1	gamma-glutamyltransferase 1	22	0.002396521
KRTAP15-1	keratin associated protein 15-1	21	0.002396521
C7orf19	-	7	0.002396521
FLJ22471	-	12	0.002396521
HK1	hexokinase 1	10	0.002505553
ADORA3	adenosine A3 receptor	-	0.002507667
PDILT	-	16	0.002509471
PCDHGA12	protocadherin gamma subfamily A, 12	5	0.00251383
MGC40222	-	6	0.00251383
FLJ22709	-	19	0.002530686
PCSK1	proprotein convertase subtilisin/kexin type 1	5	0.002692681
FBXL22	F-box and leucine-rich repeat protein 22	15	0.002724691
H2AFY	H2A histone family, member Y	5	0.002779813
TRIM2	tripartite motif-containing 2	4	0.002817627
ZFP41	zinc finger protein 41 homolog (mouse)	8	0.002826726

DYRK3	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 3	1	0.00286757
PAX4	paired box gene 4	7	0.00286757
ATP8A2	ATPase, aminophospholipid transporter-like, Class I, type 8A, member 2	13	0.00286757
SLC17A2	solute carrier family 17 (sodium phosphate), member 2	6	0.00286757
PLA2G4E	phospholipase A2, group IVE	15	0.00286757
C14orf166B	chromosome 14 open reading frame 166B	14	0.00286757
OSR1	odd-skipped related 1 (Drosophila)	2	0.003011307
PWCR1	Prader-Willi syndrome chromosome region 1	15	0.003094661
CARD9	caspase recruitment domain family, member 9	9	0.003190814
BACE2	beta-site APP-cleaving enzyme 2	21	0.003264819
ZNF677	zinc finger protein 677	19	0.00328739
GPR55	G protein-coupled receptor 55	2	0.003345195
RAB11FIP2	RAB11 family interacting protein 2 (class I)	10	0.003345195
ARPP-21	-	-	0.003345195
FABP1	fatty acid binding protein 1, liver	2	0.003360546
TNIP3	TNFAIP3 interacting protein 3	4	0.003360546
OTUD6A	OTU domain containing 6A	X	0.003526384
SULT1B1	sulfotransferase family, cytosolic, 1B, member 1	4	0.003575538
PPGB	protective protein for beta-galactosidase (galactosialidosis)	20	0.003687649
BDH	-	3	0.003689823
S100A3	S100 calcium binding protein A3	1	0.003689823
KIAA0753	KIAA0753	17	0.003689823
SCNM1	sodium channel modifier 1	1	0.003689823
ZNF177	zinc finger protein 177	19	0.003689823
PEX11G	peroxisomal biogenesis factor 11 gamma	19	0.003694372
WT1	Wilms tumor 1	11	0.003795811
C5orf13	chromosome 5 open reading frame 13	5	0.003865691
EPDR1	ependymin related protein 1 (zebrafish)	7	0.003904573
LIMS2	LIM and senescent cell antigen-like domains 2	2	0.004127572
CKMT2	creatine kinase, mitochondrial 2 (sarcomeric)	5	0.00414173
DRP2	dystrophin related protein 2	X	0.00414173
ZNF96	zinc finger protein 96	6	0.00414173
C18orf20	chromosome 18 open reading frame 20	18	0.00414173
ZIC1	Zic family member 1 (odd-paired homolog, Drosophila)	3	0.004149229
PLCE1	phospholipase C, epsilon 1	10	0.004175697
KCTD12	potassium channel tetramerisation domain containing 12	13	0.004175697
HNRPDL	heterogeneous nuclear ribonucleoprotein D-like	4	0.004175697
KCNS1	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 1	20	0.004175697
HNF4A	hepatocyte nuclear factor 4, alpha	20	0.004256414
CLEC4F	C-type lectin domain family 4, member F	2	0.004299473
ATP10A	ATPase, Class V, type 10A	15	0.004299473
GATA4	GATA binding protein 4	8	0.004311008
DNASE2B	deoxyribonuclease II beta	1	0.004311008

TAC3	tachykinin 3 (neuromedin K, neurokinin beta)	12	0.004320452
ZNF553	zinc finger protein 553	16	0.004320452
KRTAP19-2	keratin associated protein 19-2	21	0.004320452
FCN2	ficolin (collagen/fibrinogen domain containing lectin) 2 (hucolin)	9	0.004320452
RALGPS2	Ral GEF with PH domain and SH3 binding motif 2	1	0.004329614
CRYBB3	crystallin, beta B3	22	0.004338542
ITPR2	inositol 1,4,5-triphosphate receptor, type 2	12	0.004440709
SLC26A8	solute carrier family 26, member 8	6	0.004477576
ZNF154	zinc finger protein 154	19	0.004488773
EDNRB	endothelin receptor type B	13	0.005013812
GALNTL5	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 5	7	0.005013812
LOC161247	-	14	0.005013812
PABPC5	poly(A) binding protein, cytoplasmic 5	X	0.005013812
GOLGA2L1	golgi autoantigen, golgin subfamily a, 2-like 1	12	0.005013812
GABRP	gamma-aminobutyric acid (GABA) A receptor, pi	5	0.00502773
IL17R	-	22	0.005070832
LMLN	leishmanolysin-like (metallopeptidase M8 family)	3	0.005070832
KIAA1622	KIAA1622	14	0.005151934
GPR24	-	22	0.005193852
TSPAN16	tetraspanin 16	19	0.005198224
CCDC70	coiled-coil domain containing 70	13	0.005198224
FCN1	ficolin (collagen/fibrinogen domain containing) 1	9	0.005198224
B3GALT5	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 5	21	0.00523292
CXorf20	chromosome X open reading frame 20	X	0.005343204
CASP10	caspase 10, apoptosis-related cysteine peptidase	2	0.005363216
ABCG4	ATP-binding cassette, sub-family G (WHITE), member 4	-	0.005363216
CASP8	caspase 8, apoptosis-related cysteine peptidase	2	0.005363216
DKFZP586H2123	-	11	0.005363216
PRKAR1B	protein kinase, cAMP-dependent, regulatory, type I, beta	7	0.005363216
C3orf57	chromosome 3 open reading frame 57	3	0.005363216
CCRK	cell cycle related kinase	9	0.005363216
ITR	-	13	0.005363216
PLUNC	palate, lung and nasal epithelium carcinoma associated	20	0.005395921
MPPED2	metallophosphoesterase domain containing 2	11	0.005395921
IL17	-	6	0.005395921
TNFRSF10C	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain	8	0.005423356
ACPT	acid phosphatase, testicular	-	0.005513466
PLCE1	phospholipase C, epsilon 1	10	0.005540682
GPHA2	glycoprotein hormone alpha 2	11	0.005540682
TAL1	T-cell acute lymphocytic leukemia 1	1	0.005540682
TBC1D14	TBC1 domain family, member 14	4	0.005540682
FAHD2A	fumarylacetoacetate hydrolase domain containing 2A	2	0.005540682
BCHE	butyrylcholinesterase	3	0.005540682

DNALI1	dynein, axonemal, light intermediate chain 1	1	0.005540682
DDAH2	dimethylarginine dimethylaminohydrolase 2	6	0.005540682
CPNE6	copine VI (neuronal)	14	0.005540682
CHRFAM7A	CHRNA7 (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and FAM7A (family with sequence similarity 7A, exons A-E) fusion	15	0.005540682
PPAP2B	phosphatidic acid phosphatase type 2B	1	0.005540682
FGD4	FYVE, RhoGEF and PH domain containing 4	12	0.005540682
FFAR2	free fatty acid receptor 2	19	0.005564938
GSN	gelsolin (amyloidosis, Finnish type)	9	0.00562665
LAMA4	laminin, alpha 4	6	0.00562665
OR2F1	olfactory receptor, family 2, subfamily F, member 1	7	0.005638788
MACF1	microtubule-actin crosslinking factor 1	1	0.005691612
SORBS2	sorbin and SH3 domain containing 2	4	0.00586338
C1QTNF5	C1q and tumor necrosis factor related protein 5	11	0.005873138
LAT2	linker for activation of T cells family, member 2	7	0.005873138
DHRS10	dehydrogenase/reductase (SDR family) member 10	19	0.005874618
CCL8	chemokine (C-C motif) ligand 8	17	0.005874618
COL8A2	collagen, type VIII, alpha 2	1	0.005874618
TANK	TRAF family member-associated NFKB activator	2	0.005874618
KRTAP8-1	keratin associated protein 8-1	21	0.005874618
ARPP-21	-	-	0.005891195
SCUBE3	signal peptide, CUB domain, EGF-like 3	6	0.005891195
FLJ22555	-	2	0.005896161
FAM20B	family with sequence similarity 20, member B	1	0.005926756
FCGBP	Fc fragment of IgG binding protein	19	0.00597756
MFSD7	major facilitator superfamily domain containing 7	4	0.005984572
SORBS3	sorbin and SH3 domain containing 3	8	0.005984572
NANOS1	nanos homolog 1 (Drosophila)	10	0.005995257
G0S2	G0/G1switch 2	1	0.006090998
PCDH8	protocadherin 8	13	0.006122297
SCN7A	sodium channel, voltage-gated, type VII, alpha	2	0.006122297
SRI	sorcin	7	0.006253498
P8	-	16	0.006253498
KLK4	kallikrein 4 (prostase, enamel matrix, prostate)	19	0.006255945
AQP2	aquaporin 2 (collecting duct)	-	0.006274651
DCUN1D1	DCN1, defective in cullin neddylation 1, domain containing 1 (S. cerevisiae)	3	0.006411689
C12orf46	chromosome 12 open reading frame 46	12	0.006411689
MT1H	metallothionein 1H	16	0.006417031
BRCA1	breast cancer 1, early onset	17	0.006417031
-	-	-	0.006447459
DOCK2	dedicator of cytokinesis 2	5	0.006447459
GLRA1	glycine receptor, alpha 1 (startle disease/hyperekplexia, stiff man syndrome)	5	0.006447459
CDH9	cadherin 9, type 2 (T1-cadherin)	5	0.006469237
STAT5B	signal transducer and activator of transcription 5B	17	0.006469237

CDCP2	CUB domain containing protein 2	1	0.006469237
SCGB1D2	secretoglobin, family 1D, member 2	11	0.006512801
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isoform	12	0.006517318
CETN1	centrin, EF-hand protein, 1	18	0.006568288
CCL13	chemokine (C-C motif) ligand 13	17	0.006598422
TBC1D3	TBC1 domain family, member 3	17	0.00661871
PDGFRA	platelet-derived growth factor receptor, alpha polypeptide	4	0.006627551
CCR3	chemokine (C-C motif) receptor 3	3	0.006650815
C19orf19	chromosome 19 open reading frame 19	19	0.006666293
TWIST1	twist homolog 1 (acrocephalosyndactyly 3	7	0.006741725
H2AFY	H2A histone family, member Y	5	0.006741725
SLC22A18	solute carrier family 22 (organic cation transporter), member 18	11	0.006741725
SNRPN	small nuclear ribonucleoprotein polypeptide N	15	0.006741725
MGC7036	-	12	0.006741725
Kua	-	20	0.006756329
MAGEC1	melanoma antigen family C, 1	X	0.006797953
ATP6V0C	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c	16	0.006825396
MSR1	macrophage scavenger receptor 1	8	0.006873407
CAB39L	calcium binding protein 39-like	13	0.006873407
ACTA2	actin, alpha 2, smooth muscle, aorta	-	0.006898194
GGTLA4	gamma-glutamyltransferase-like activity 4	20	0.006923809
TMEM88	transmembrane protein 88	17	0.007011951
AASS	aminoadipate-semialdehyde synthase	-	0.007011951
GDF5	growth differentiation factor 5 (cartilage-derived morphogenetic protein-1)	20	0.007011951
IVL	involucrin	1	0.007046375
TLR4	toll-like receptor 4	9	0.007147819
ME3	malic enzyme 3, NADP(+)-dependent, mitochondrial	11	0.007147819
GRIA2	glutamate receptor, ionotropic, AMPA 2	4	0.007162973
DYSF	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)	2	0.007203188
TRIM60	tripartite motif-containing 60	4	0.007215958
ZNF80	zinc finger protein 80	3	0.007299624
TBC1D3	TBC1 domain family, member 3	17	0.007395639
-	-	-	0.007421584
ATM	ataxia telangiectasia mutated (includes complementation groups A, C and D)	11	0.007421584
RNH1	ribonuclease/angiogenin inhibitor 1	11	0.007421584
HDAC1	histone deacetylase 1	1	0.007496188
CST4	cystatin S	20	0.007496188
MRGPRX4	MAS-related GPR, member X4	11	0.007496188
FLJ36004	-	12	0.00750587
ZNF659	zinc finger protein 659	3	0.00750587
CLC	Charcot-Leyden crystal protein	19	0.00751797
ZNF667	zinc finger protein 667	19	0.007540398
ZNF454	zinc finger protein 454	5	0.007550939

DSG1	desmoglein 1	18	0.007618564
NA	neurocanthocytosis	-	0.007618564
ABCC10	ATP-binding cassette, sub-family C (CFTR/MRP), member 10	6	0.007821999
CSNK1A1L	casein kinase 1, alpha 1-like	13	0.007860271
SLC5A4	solute carrier family 5 (low affinity glucose cotransporter), member 4	22	0.007939866
RLN3R1	-	5	0.007949721
CDK2	cyclin-dependent kinase 2	12	0.00825271
POU2F2	POU domain, class 2, transcription factor 2	19	0.008391955
ITGB1BP2	integrin beta 1 binding protein (melusin) 2	X	0.008391955
FAM9B	family with sequence similarity 9, member B	X	0.008391955
GCM2	glial cells missing homolog 2 (Drosophila)	6	0.008400027
FLJ23514	-	11	0.008400027
MRAS	muscle RAS oncogene homolog	3	0.008411015
SARDH	sarcosine dehydrogenase	9	0.008497573
SLC6A2	solute carrier family 6 (neurotransmitter transporter, noradrenalin), member 2	16	0.00853534
ORMDL3	ORM1-like 3 (S. cerevisiae)	17	0.008660065
WT1	Wilms tumor 1	11	0.008660065
MGC40178	-	13	0.008660065
SLC22A18AS	solute carrier family 22 (organic cation transporter), member 18 antisense	11	0.008660065
PCDHGB4	protocadherin gamma subfamily B, 4	5	0.008717794
MAEL	maelstrom homolog (Drosophila)	1	0.008717794
MEGF10	multiple EGF-like-domains 10	5	0.008721123
FAM9A	family with sequence similarity 9, member A	X	0.008755454
STEAP2	six transmembrane epithelial antigen of the prostate 2	7	0.008755454
GALNT5	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5)	2	0.008828744
DNAJB8	DnaJ (Hsp40) homolog, subfamily B, member 8	3	0.008830444
ADH7	alcohol dehydrogenase 7 (class IV), mu or sigma polypeptide	4	0.008893124
MS4A7	membrane-spanning 4-domains, subfamily A, member 7	11	0.008893124
FMR1NB	fragile X mental retardation 1 neighbor	X	0.008895268
XAGE5	X antigen family, member 5	X	0.008902474
CCL7	chemokine (C-C motif) ligand 7	17	0.008930512
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	9	0.008979211
UGT2B4	UDP glucuronosyltransferase 2 family, polypeptide B4	4	0.009015415
BIRC8	baculoviral IAP repeat-containing 8	19	0.009015415
REG1A	regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)	2	0.009015415
NRTN	neurturin	19	0.009015415
RNF2	ring finger protein 2	1	0.009015415
IRS1	insulin receptor substrate 1	2	0.009015415
KIF5A	kinesin family member 5A	12	0.009015415
TTLL6	tubulin tyrosine ligase-like family, member 6	17	0.009015415
SNRK	SNF related kinase	3	0.009015415

EPAS1	endothelial PAS domain protein 1	2	0.009015415
JM11	-	X	0.00901593
DCC	deleted in colorectal carcinoma	18	0.009175824
TMCO5	transmembrane and coiled-coil domains 5	15	0.009187483
ZNF76	zinc finger protein 76 (expressed in testis)	6	0.009259825
INS	insulin	11	0.009384921
COL11A2	collagen, type XI, alpha 2	6	0.009384921
REG3A	regenerating islet-derived 3 alpha	2	0.009384921
CIB3	calcium and integrin binding family member 3	19	0.009384921
LYZL6	lysozyme-like 6	17	0.009467574
HIST1H2AA	histone cluster 1, H2aa	6	0.009467574
ANXA13	annexin A13	8	0.009467574
CD5L	CD5 molecule-like	1	0.009467574
SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	10	0.009476994
KRTAP1-1	keratin associated protein 1-1	17	0.009574444
MYH1	myosin, heavy chain 1, skeletal muscle, adult	17	0.009574444
ENDOGL1	endonuclease G-like 1	3	0.009574444
C10orf111	chromosome 10 open reading frame 111	10	0.009582908
CSDC2	cold shock domain containing C2, RNA binding	22	0.009615483
PHYHIP	phytanoyl-CoA 2-hydroxylase interacting protein	8	0.009615483
BCDIN3	bin3, bicoid-interacting 3, homolog (Drosophila)	7	0.009759351
FHL5	four and a half LIM domains 5	6	0.009759351
KCNE1	potassium voltage-gated channel, Isk-related family, member 1	21	0.00984379
DPM2	dolichyl-phosphate mannosyltransferase polypeptide 2, regulatory subunit	9	0.009921441

Appendix 8. List of genes from the previous appendix shown to be involved in cancer based on literature search

Gene	Prevalent Function based on literature search	Methylation Status in cSCC cell lines
SPARCL1	Tumour Suppressor	Hyper
CDKN2A	Tumour Suppressor	Hyper
RGS17	Oncogene	Hyper
FGFR1	Oncogene	Hyper
BARHL2	Tumour Suppressor	Hyper
CASP8	Pro-apoptotic	Hypo
NEFH	Tumour Suppressor	Hyper
TMEM22	Oncogene	Hyper
PRKDC	Oncogene	Hypo
ZDHHC11	Oncogene	Hypo
CR1	Tumour Suppressor	Hyper
NDRG2	Tumour Suppressor	Hyper
TGIF2	Acts as both TS and oncogene	Hyper
PRAME	Oncogene	Hypo
TP53I3	Tumour Suppressor	Hyper
KRT7	Tumour Suppressor	Hyper
HSPCAL3	Oncogene	Hypo
SERPINA3	Tumour Suppressor	Hyper
NRTN	Acts as both TS and oncogene	Hypo
B4GALT1	Tumour Suppressor	Hyper
LSAMP	Tumour Suppressor	Hyper
UCHL1	Tumour Suppressor	Hyper
TBC1D3C	Oncogene	Hypo
THSD1	Tumour Suppressor	Hypo
GAGE4	Oncogene	Hypo
MMP20	Oncogene	Hypo
C9orf78	Tumour Suppressor	Hyper
ITK	Oncogene	Hypo
PRKCB1	Acts as both TS and oncogene	Hyper
CCT6A	Oncogene	Hypo
SIM1	Tumour Suppressor	Hyper
PARC	Tumour Suppressor	Hyper
SST	Acts as both TS and oncogene	Hyper
ZIC1	Acts as both TS and oncogene	Hyper
CNN1	Tumour Suppressor	Hyper
TNFRSF10C	Tumour Suppressor	Hyper
TNFRSF1B	Oncogene	Hyper
SMPD3	Tumour Suppressor	Hyper
TBC1D3C	Oncogene	Hypo

GATA4	Tumour Suppressor	Hyper
SPINK2	Oncogene	Hyper
REGL	Oncogene	Hypo
AQP9	Acts as both TS and oncogene	Hyper
GGT1	Oncogene	Hyper
HK1	Oncogene	Hypo
ADORA3	Oncogene	Hypo
H2AFY	Tumour Suppressor	Hyper
DYRK3	Oncogene	Hyper
PAX4	Acts as both TS and oncogene	Hypo
GPR55	Oncogene	Hypo
FABP1	Oncogene	Hypo
S100A3	Acts as both TS and oncogene	Hyper
WT1	Tumour Suppressor	Hyper
EPDR1	Oncogene	Hypo
LIMS2	Tumour Suppressor	Hypo
PLCE1	Oncogene	Hypo
KCTD12	Tumour Suppressor	Hyper
HNF4A	Oncogene	Hypo
EDNRB	Acts as both TS and oncogene	Hypo
GABRP	Tumour Suppressor	Hypo
TSPAN16	Tumour Suppressor	Hyper
B3GALT5	Oncogene	Hypo
CCRK	Oncogene	Hyper
PLUNC	Oncogene	Hypo
IL17	Oncogene	Hypo
TAL1	Tumour Suppressor	Hyper
BCHE	Acts as both TS and oncogene	Hyper
CPNE6	Oncogene	Hypo
GSN	Tumour Suppressor	Hyper
LAMA4	Acts as both TS and oncogene	Hyper
SORBS2	Tumour Suppressor	Hyper
LAT2	Oncogene	Hypo
SORBS3	Oncogene	Hyper
NANOS1	Oncogene	Hyper
G0S2	Tumour Suppressor	Hyper
PCDH8	Tumour Suppressor	Hyper
SRI	Oncogene	Hypo
P8	Tumour Suppressor	Hyper
KLK4	Oncogene	Hypo
DCUN1D1	Oncogene	Hyper
MT1H	Oncogene	Hypo
BRCA1	Tumour Suppressor	Hypo
DOCK2	Oncogene	Hyper
STAT5B	Tumour Suppressor	Hyper
SCGB1D2	Oncogene	Hypo
PDGFRA	Oncogene	Hyper
CCR3	Oncogene	Hypo
TWIST1	Tumour Suppressor	Hyper
SLC22A18	Tumour Suppressor	Hyper
MAGEC1	Oncogene	Hypo

MSR1	Acts as both TS and oncogene	Hypo
GDF5	Acts as both TS and oncogene	Hypo
IVL	Acts as both TS and oncogene	Hyper
TLR4	Oncogene	Hypo
TBC1D3	Oncogene	Hypo
ATM	Tumour Suppressor	Hypo
HDAC1	Oncogene	Hypo
DSG1	Tumour Suppressor	Hyper
RLN3R1	Acts as both TS and oncogene	Hyper
CDK2	Acts as both TS and oncogene	Hyper
POU2F2	Acts as both TS and oncogene	Hyper
MRAS	Oncogene	Hypo
SLC6A2	Tumour Suppressor	Hyper
MAEL	Oncogene	Hypo
STEAP2	Oncogene	Hyper
ADH7	Acts as both TS and oncogene	Hypo
CCL7	Oncogene	Hypo
BIRC8	Oncogene	Hypo
REG1A	Oncogene	Hypo
RNF2	Oncogene	Hyper
IRS1	Acts as both TS and oncogene	Hyper
DCC	Acts as both TS and oncogene	Hypo
REG3A	Oncogene	Hypo
ANXA13	Acts as both TS and oncogene	Hyper
SPOCK2	Tumour Suppressor	Hyper

Appendix 9. List of 76 miRNAs differentially expressed in cSCC compared to normal skin.

Name	logFC	AveExpr	P.Value	adj.P.Val
hsa-miR-1469	-1.072505836	8.909492183	5.77436154519119e-10	1.20972874371755e-06
hsv1-miR-H17	-0.705890727	6.293762043	1.69129214680814e-09	1.77162852378153e-06
hsa-miR-548ap-5p/hsa-miR-548j	-1.137583927	6.387216938	6.37345196612178e-09	4.45079395634171e-06
hsa-miR-3656	-1.010574316	7.992558753	2.62424656124373e-08	1.3744491364514e-05
hsa-miR-4707-5p	-0.899529655	7.154120387	3.54495753467037e-08	1.48533720702689e-05
hsa-miR-371b-5p	-0.796315305	12.95202837	7.02535328882583e-08	2.45301919001502e-05
hsa-miR-638	-0.96944656	7.528746852	8.79507138186871e-08	2.63223922071642e-05
hsa-miR-4787-5p	-0.783080099	13.39009682	1.0748940076033e-07	2.81487868241115e-05
hsa-miR-4708-3p	-0.819929533	13.73263792	4.15585920532755e-07	9.6739167057347e-05
hsa-miR-4530	-0.912825789	9.016009311	5.83423681553376e-07	0.000122227
hsa-miR-4800-3p	-0.959652136	12.58479102	8.73797238933443e-07	0.000166419
hsa-miR-943	-0.547735401	7.141100883	1.53754766534614e-06	0.00026843
kshv-miR-K12-3-5p	-0.582550336	7.324808865	1.99976954076941e-06	0.00030663
hsa-miR-3960	-0.745469234	13.55820195	2.04908076517478e-06	0.00030663
hsa-miR-1909-3p	-0.625525465	6.088837698	2.71122173072946e-06	0.000378667
hsa-miR-4505	-0.965317513	8.646416497	7.52660351418713e-06	0.000985515
hsa-miR-744-5p	-0.691696164	7.964567669	9.56746704377922e-06	0.00117905
hsa-miR-642b-3p	1.141917999	6.608054895	1.15649194007355e-05	0.001275185
hsa-miR-605	-0.513822616	5.813879982	1.15575701021536e-05	0.001275185

hsa-miR-31-5p	1.740874987	6.468657902	1.25833940732504e-05	0.001318111
hsa-miR-21-3p	0.6724371	5.104375825	1.42998835672022e-05	0.001426584
hsa-miR-4732-5p	1.92248161	9.252727944	2.03475247609164e-05	0.001937639
hsa-miR-4423-5p	0.443625999	5.250061222	3.69901561629572e-05	0.003228932
kshv-miR-K12-6-3p	-0.793278205	7.928657316	3.57816180117349e-05	0.003228932
hsa-miR-585	-0.617744304	5.599632618	3.90871812017574e-05	0.003275506
hsa-miR-125b-1-3p	-0.349733408	5.621494536	4.82440516267739e-05	0.003609689
hsa-miR-3195	-0.694960429	6.110821849	4.49031145967249e-05	0.003609689
hsv1-miR-H7-3p	-0.708660707	9.884600092	4.66662449625971e-05	0.003609689
hsa-miR-495-5p	0.391768945	5.66288775	6.69281365222725e-05	0.004523047
hsv2-miR-H10	-0.713452548	8.522406314	6.68499148385216e-05	0.004523047
hsa-miR-663a	-0.749102069	7.506702693	6.44310090858787e-05	0.004523047
hsa-miR-3687	-0.865385123	7.2448325	8.8350480273067e-05	0.005784196
hsa-miR-4516	-0.697507752	10.99051286	0.000101523	0.006445202
hsa-miR-4734	-0.445995207	5.272286218	0.000112973	0.006961121
hsa-miR-124-5p	-0.298243171	6.10332581	0.000127269	0.007617956
hsa-miR-224-3p	-0.561808414	6.789616544	0.00013696	0.007970294
hsa-miR-1290	1.530172368	8.672801553	0.000147766	0.008366759
hsa-miR-639	-0.249764274	5.584202327	0.000178618	0.009847496
hsa-miR-3182	2.337082252	8.163444664	0.000217911	0.011413106
hsa-miR-4488	-0.454136103	6.913694905	0.00021514	0.011413106
hsa-miR-4467	-0.547369411	12.57858681	0.00022758	0.011462915
hsa-miR-149-3p	-0.729607545	7.543904852	0.000229805	0.011462915
hsa-miR-193b-3p	1.118807996	7.5479942	0.000335596	0.015978967

hsa-miR-766-3p	0.207156908	5.142699547	0.000333664	0.015978967
hsa-miR-498	-0.306547631	6.769322106	0.000430022	0.0200199
hsa-miR-3920	0.216338628	4.840038311	0.000473974	0.021586409
hsa-miR-3935	0.43234109	6.381786768	0.000521247	0.022821632
hsa-miR-130b-3p	0.293904195	5.765020465	0.000531654	0.022821632
hsa-miR-4750-5p	-0.59968878	9.784396476	0.000533776	0.022821632
hsa-miR-24-2-5p	0.365208776	5.027875459	0.000548958	0.02300134
hsa-miR-4447	-0.273777285	5.626449958	0.000580342	0.023839532
hsa-miR-3156-3p	0.395539349	6.820620878	0.000681503	0.027456701
hsa-miR-135b-5p	0.883937247	5.037672863	0.000851169	0.033645257
hsv2-miR-H7-3p	-0.589493933	9.321282767	0.000899884	0.034912168
hsa-miR-21-5p	2.246588244	8.033722502	0.000934032	0.03494282
hsa-miR-423-3p	0.579440254	6.524077005	0.000920679	0.03494282
hsa-miR-223-3p	1.763572745	7.236146792	0.001020281	0.036254584
hsa-miR-1307-5p	0.285578386	4.757912697	0.001021012	0.036254584
hsa-miR-4508	-0.98043825	8.831630896	0.001010704	0.036254584
hsa-miR-1246	1.419004492	13.60041948	0.001045013	0.03648836
hsv1-miR-H5-3p	0.283271184	5.040945818	0.001191264	0.040673136
hsa-miR-1908	-0.467009308	10.32439581	0.001203692	0.040673136
hsa-miR-4521	0.621576095	5.573759197	0.001249158	0.041539459
hsa-miR-4286	0.956050588	7.537456837	0.00127756	0.041820135
hsa-miR-5581-3p	0.599733688	6.679426255	0.001351191	0.043549929
hsa-miR-3676-3p	-0.343569654	7.661996726	0.001404249	0.044574273
hsa-miR-210	0.71365027	7.068375043	0.001464654	0.045797756
hsa-miR-142-5p	0.99340117	5.087023745	0.001600617	0.047424561

hsa-miR-4732-3p	0.654731845	7.779519857	0.00156408	0.047424561
hsa-miR-1255b-2-3p	0.374734845	5.193972354	0.001588268	0.047424561
hsa-let-7d-3p	-0.3890342	6.9052577	0.001607229	0.047424561
hsa-miR-4639-3p	0.460282559	9.03155916	0.001691256	0.048200154
hsa-miR-4320	0.217947719	4.764371146	0.001688837	0.048200154
hcmv-miR-US25-1-3p	-0.251550024	5.384550166	0.001702535	0.048200154
hsa-miR-93-5p	0.990397389	6.667981642	0.001768337	0.049395534
hsa-miR-4317	0.950932916	6.013905017	0.001798702	0.049582645

Appendix 10. KEGG Pathway analysis of genes hypomethylated in AK.

KEGG Pathway	Annotated genes	p-value	Adjusted p value
Neuroactive ligand-receptor interaction	44	6.45E-16	1.10E-13
Axon guidance	30	1.18E-15	1.10E-13
MAPK signaling pathway	41	3.84E-14	2.38E-12
Focal adhesion	34	3.35E-13	1.56E-11
Calcium signaling pathway	29	4.60E-11	1.71E-09
Type II diabetes mellitus	15	1.21E-10	3.74E-09
Hypertrophic cardiomyopathy (HCM)	19	4.01E-10	1.06E-08
Pathways in cancer	39	4.56E-10	1.06E-08
Dilated cardiomyopathy	18	1.10E-08	2.28E-07
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	16	2.39E-08	4.45E-07
Tight junction	21	4.20E-08	7.10E-07
ErbB signaling pathway	16	1.76E-07	2.73E-06
Cell adhesion molecules (CAMs)	20	1.99E-07	2.85E-06
Regulation of actin cytoskeleton	26	2.70E-07	3.59E-06
Leukocyte transendothelial migration	18	5.75E-07	7.13E-06
Melanogenesis	16	1.64E-06	1.91E-05
Vascular smooth muscle contraction	17	1.84E-06	2.01E-05
Long-term depression	13	2.20E-06	2.28E-05
Purine metabolism	20	3.15E-06	3.08E-05
ECM-receptor interaction	14	3.48E-06	3.23E-05
Insulin signaling pathway	18	5.19E-06	4.60E-05
Olfactory transduction	34	7.99E-06	6.75E-05
Long-term potentiation	12	1.27E-05	1.03E-04
Aldosterone-regulated sodium reabsorption	9	2.42E-05	1.88E-04
Glioma	11	3.29E-05	2.45E-04
Fc epsilon RI signaling pathway	12	4.46E-05	3.19E-04
Cardiac muscle contraction	12	5.07E-05	3.49E-04
Histidine metabolism	7	8.75E-05	5.81E-04
Drug metabolism - other enzymes	9	1.21E-04	7.78E-04
Adherens junction	11	1.27E-04	7.88E-04
Gap junction	12	1.62E-04	9.74E-04
Adipocytokine signaling pathway	10	2.20E-04	1.28E-03
Wnt signaling pathway	16	2.31E-04	1.30E-03
Pantothenate and CoA biosynthesis	5	2.49E-04	1.33E-03
T cell receptor signaling pathway	13	2.50E-04	1.33E-03
Progesterone-mediated oocyte maturation	11	4.33E-04	2.24E-03
Phenylalanine metabolism	5	4.57E-04	2.30E-03

GnRH signaling pathway	12	4.81E-04	2.35E-03
Endocytosis	17	6.99E-04	3.33E-03
mTOR signaling pathway	8	7.52E-04	3.50E-03
Natural killer cell mediated cytotoxicity	14	7.95E-04	3.61E-03
O-Glycan biosynthesis	6	8.32E-04	3.68E-03
Amyotrophic lateral sclerosis (ALS)	8	8.57E-04	3.71E-03
Tyrosine metabolism	7	9.80E-04	4.14E-03
Chemokine signaling pathway	17	1.06E-03	4.39E-03
Neurotrophin signaling pathway	13	1.10E-03	4.46E-03
Fc gamma R-mediated phagocytosis	11	1.20E-03	4.76E-03
ABC transporters	7	1.30E-03	4.95E-03
Amino sugar and nucleotide sugar metabolism	7	1.30E-03	4.95E-03
Small cell lung cancer	10	1.37E-03	5.08E-03
Proximal tubule bicarbonate reclamation	5	1.53E-03	5.59E-03
Fructose and mannose metabolism	6	1.65E-03	5.92E-03
Ribosome	10	1.95E-03	6.84E-03
Glycolysis / Gluconeogenesis	8	2.42E-03	8.20E-03
Phosphatidylinositol signaling system	9	2.42E-03	8.20E-03
Peroxisome	9	2.90E-03	9.63E-03
SNARE interactions in vesicular transport	6	2.98E-03	9.73E-03
Endometrial cancer	7	3.49E-03	1.12E-02
Cell cycle	12	3.75E-03	1.18E-02
Non-small cell lung cancer	7	4.33E-03	1.34E-02
Pancreatic cancer	8	5.16E-03	1.55E-02
Renal cell carcinoma	8	5.16E-03	1.55E-02
Hedgehog signaling pathway	7	5.31E-03	1.57E-02
TGF-beta signaling pathway	9	5.58E-03	1.61E-02
Melanoma	8	5.63E-03	1.61E-02
Type I diabetes mellitus	6	6.28E-03	1.77E-02
Chronic myeloid leukemia	8	6.66E-03	1.85E-02
Prostate cancer	9	6.98E-03	1.91E-02
Pathogenic Escherichia coli infection	7	7.08E-03	1.91E-02
Acute myeloid leukemia	7	7.75E-03	2.06E-02
Notch signaling pathway	6	8.68E-03	2.27E-02
Oocyte meiosis	10	1.21E-02	3.14E-02
Dorso-ventral axis formation	4	1.41E-02	3.61E-02
Cytokine-cytokine receptor interaction	18	1.49E-02	3.74E-02
Complement and coagulation cascades	7	1.61E-02	3.97E-02
Galactose metabolism	4	1.62E-02	3.97E-02
Pyruvate metabolism	5	1.76E-02	4.24E-02
Basal cell carcinoma	6	1.81E-02	4.32E-02
Systemic lupus erythematosus	11	1.88E-02	4.43E-02
Apoptosis	8	1.94E-02	4.45E-02
Hematopoietic cell lineage	8	1.94E-02	4.45E-02
Leishmania infection	7	2.00E-02	4.54E-02
Graft-versus-host disease	5	2.13E-02	4.70E-02

Glyoxylate and dicarboxylate metabolism	3	2.15E-02	4.70E-02
Riboflavin metabolism	3	2.15E-02	4.70E-02

Appendix 11. KEGG Pathway analysis of genes hypermethylated in AK.

Description	Genes	p-value	Adjusted p value
Pathways in cancer	56	1.03E-15	1.92E-13
Wnt signaling pathway	35	2.19E-14	2.03E-12
Focal adhesion	39	3.45E-13	2.14E-11
Endocytosis	35	8.44E-12	3.93E-10
PPAR signaling pathway	20	1.04E-10	3.88E-09
Glycerophospholipid metabolism	21	1.27E-10	3.95E-09
Jak-STAT signaling pathway	29	8.58E-10	2.28E-08
Melanogenesis	23	1.06E-09	2.47E-08
Phosphatidylinositol signaling system	19	4.74E-09	9.79E-08
Calcium signaling pathway	29	2.26E-08	4.11E-07
MAPK signaling pathway	37	2.55E-08	4.11E-07
Axon guidance	24	2.65E-08	4.11E-07
Basal cell carcinoma	15	5.54E-08	7.85E-07
Vascular smooth muscle contraction	22	5.91E-08	7.85E-07
Prostate cancer	19	7.50E-08	9.31E-07
Adherens junction	17	1.43E-07	1.65E-06
Regulation of actin cytoskeleton	31	1.51E-07	1.65E-06
Chemokine signaling pathway	28	3.40E-07	3.50E-06
Insulin signaling pathway	23	3.58E-07	3.50E-06
Glycosaminoglycan biosynthesis - chondroitin sulfate	9	5.50E-07	5.12E-06
GnRH signaling pathway	19	6.07E-07	5.37E-06
ECM-receptor interaction	17	8.02E-07	6.78E-06
Non-small cell lung cancer	13	2.01E-06	1.62E-05
Notch signaling pathway	12	2.54E-06	1.97E-05
Dilated cardiomyopathy	17	3.02E-06	2.23E-05
Hedgehog signaling pathway	13	3.12E-06	2.23E-05
Small cell lung cancer	16	3.88E-06	2.67E-05
Glycerolipid metabolism	12	4.08E-06	2.71E-05
ErbB signaling pathway	16	6.24E-06	4.00E-05
Acute myeloid leukemia	13	7.07E-06	4.38E-05
Vasopressin-regulated water reabsorption	11	8.28E-06	4.97E-05
Colorectal cancer	13	1.03E-05	6.01E-05
Tight junction	20	1.27E-05	7.16E-05
Chronic myeloid leukemia	14	1.42E-05	7.75E-05
Oocyte meiosis	18	1.55E-05	8.24E-05
Neurotrophin signaling pathway	19	1.79E-05	9.24E-05
Neuroactive ligand-receptor interaction	31	2.03E-05	1.02E-04

Cell cycle	19	2.24E-05	1.08E-04
Progesterone-mediated oocyte maturation	15	2.33E-05	1.08E-04
TGF-beta signaling pathway	15	2.33E-05	1.08E-04
Adipocytokine signaling pathway	13	2.50E-05	1.13E-04
Fc gamma R-mediated phagocytosis	16	2.60E-05	1.15E-04
Pancreatic cancer	13	4.06E-05	1.76E-04
Endometrial cancer	11	4.50E-05	1.86E-04
mTOR signaling pathway	11	4.50E-05	1.86E-04
Inositol phosphate metabolism	11	6.50E-05	2.63E-04
Hypertrophic cardiomyopathy (HCM)	14	8.29E-05	3.26E-04
Glioma	12	8.46E-05	3.26E-04
B cell receptor signaling pathway	13	8.58E-05	3.26E-04
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	13	9.88E-05	3.62E-04
T cell receptor signaling pathway	16	9.91E-05	3.62E-04
Gap junction	14	1.56E-04	5.58E-04
Long-term potentiation	12	1.77E-04	6.10E-04
Renal cell carcinoma	12	1.77E-04	6.10E-04
Circadian rhythm - mammal	5	2.92E-04	9.88E-04
Apoptosis	13	4.44E-04	1.45E-03
Hematopoietic cell lineage	13	4.44E-04	1.45E-03
Thyroid cancer	7	4.70E-04	1.51E-03
Ubiquitin mediated proteolysis	17	5.81E-04	1.83E-03
Other glycan degradation	5	8.74E-04	2.71E-03
Cytokine-cytokine receptor interaction	26	1.05E-03	3.20E-03
Huntington's disease	20	1.08E-03	3.25E-03
Lysine degradation	8	1.41E-03	4.15E-03
Fc epsilon RI signaling pathway	11	1.95E-03	5.66E-03
Melanoma	10	2.82E-03	8.07E-03
Biosynthesis of unsaturated fatty acids	5	4.10E-03	1.15E-02
VEGF signaling pathway	10	4.65E-03	1.29E-02
Ether lipid metabolism	6	5.50E-03	1.50E-02
Leukocyte transendothelial migration	13	6.44E-03	1.73E-02
Purine metabolism	16	6.52E-03	1.73E-02
Vibrio cholerae infection	8	6.63E-03	1.74E-02
Cell adhesion molecules (CAMs)	14	7.64E-03	1.97E-02
Arachidonic acid metabolism	8	8.19E-03	2.09E-02
RNA degradation	8	9.07E-03	2.28E-02
Taurine and hypotaurine metabolism	3	1.16E-02	2.89E-02
Homologous recombination	5	1.19E-02	2.92E-02
alpha-Linolenic acid metabolism	4	1.34E-02	3.23E-02
Toll-like receptor signaling pathway	11	1.35E-02	3.23E-02
Histidine metabolism	5	1.38E-02	3.25E-02
Valine, leucine and isoleucine biosynthesis	3	1.54E-02	3.59E-02
Aminoacyl-tRNA biosynthesis	6	1.59E-02	3.61E-02
O-Glycan biosynthesis	5	1.59E-02	3.61E-02

Amyotrophic lateral sclerosis (ALS)	7	1.63E-02	3.66E-02
Lysosome	12	1.88E-02	4.17E-02
p53 signaling pathway	8	2.21E-02	4.83E-02

Appendix 12. Transcription factors targeting genes with both hypermethylated and hypomethylated promoters.

Motif Shared	Transcription Factor
NKNTTGCNYAAYNN	CEBPB: CCAAT/enhancer binding protein (C/EBP), beta
AGCAHAC	DBP: D site of albumin promoter (albumin D-box) binding protein
TGACCTY	ESRRA: estrogen-related receptor alpha
RYTTCCTG	ETS2: v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
TGTTTGY	FOXA1: forkhead box A1
NNNNNGTAAATAAACA	FOXC1: forkhead box C1
RTAAACA	FOXF2: forkhead box F2
NNANNGTAAACAANNN	FOXF2: forkhead box F2
NNCWGATARNNNN	GATA1: GATA binding protein 1 (globin transcription factor 1)
NGATAAGNMNN	GATA1: GATA binding protein 1 (globin transcription factor 1)
NCWGATAACA	GATA1: GATA binding protein 1 (globin transcription factor 1)
NNGATARNG	GATA3: GATA binding protein 3
TGATTTRY	GFI1: growth factor independent 1
NNNNNNNAAATCACWGYNNNNNNN	GFI1: growth factor independent 1
TTCACTT	IRF1: interferon regulatory factor 1
TGANTCA	JUN: jun oncogene
CTTTGT	LEF1: lymphoid enhancer-binding factor 1
CTTTGA	LEF1: lymphoid enhancer-binding factor 1
AATTAATTAA	LHX3: LIM homeobox 3
GGGAGGRR	MAZ: MYC-associated zinc finger protein (purine-binding transcription factor)
YTATTTTNR	MEF2A: MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
RNKCTATTTWTAGMWN	MEF2A: MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
TGACAGNY	MEIS1: Meis1, myeloid ecotropic viral integration site 1 homolog (mouse)
TTGTTT	MLLT7: myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 7
GCANCTGNY	MYOD1: myogenic differentiation 1
TGGAAA	NFAT NFATC
GGGTGGRR	PAX4: paired box gene 4
GGATTA	PITX2: paired-like homeodomain transcription factor 2
NNNNWTATGCAAATNTNNN	POU2F1: POU domain, class 2, transcription factor 1
NNNNNNNWATGCAAATNNNWN	POU2F1: POU domain, class 2, transcription factor 1

NA	
NNCATNSRWAATNMRN	POU3F2: POU domain, class 3, transcription factor 2
GCATAAWTTAT	POU6F1: POU domain, class 6, transcription factor 1
CAGCTG	REPIN1: replication initiator 1
GGGCGGR	SP1: Sp1 transcription factor
TATAAA	TAF TATA
CAGGTG	TCF3: transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
CAGGTA	TCF8: transcription factor 8 (represses interleukin 2 expression)
TAATTA	VSX1: visual system homeobox 1 homolog, CHX10-like (zebrafish)
YTATTTTNR	MEF2A: MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
YCATTAA	Unknown
YATGNWAAT	Unknown
WTTGKCTG	Unknown
WTGAAAT	Unknown
WGTTNNNNNAAA	Unknown
TTANTCA	Unknown
TGACATY	Unknown
SMTTTTGT	Unknown
NNNNNNKCTAWAAATAGMNNNNN	Unknown
GCTNWTTGK	Unknown
GATAAGR	Unknown
CTTTAAR	Unknown
CTGCAGY	Unknown
AACTTT	Unknown

Appendix 13. Transcription factors targeting genes with hypomethylated promoters.

Motif-Hyper	Transcription Factor
TGTGGTTW	CBFA2T2: core-binding factor, runt domain, alpha subunit 2; translocated to, 2 CBFA2T3: core-binding factor, runt domain, alpha subunit 2; translocated to, 3
NNATTRCNNAANNN	CEBPA: CCAAT/enhancer binding protein (C/EBP), alpha
NNTKTGGWNANNN	CEBPA: CCAAT/enhancer binding protein (C/EBP), alpha
RNRTKDNGMAAKNN	CEBPB: CCAAT/enhancer binding protein (C/EBP), beta
CTBATTTCARAAW	CEBPG: CCAAT/enhancer binding protein (C/EBP), gamma
YNNNTAATCYCMN	CRX: cone-rod homeobox
NATYGATSSS	CUTL1: cut-like 1, CCAAT displacement protein (Drosophila)
CTWAWGTAAACANWGN	FOXD1: forkhead box D1
NNNTGTTTATNTR	FOXJ1: forkhead box J1
NNNWAAAYAAAYANNNNN	FOXJ2: forkhead box J2
AYMATAATATTTKN	FOXJ2: forkhead box J2
ARATKGAST	FOXM1: forkhead box M1
NRWAAACAAN	FOXO1A: forkhead box O1A (rhabdomyosarcoma)
NATTGTTTATWT	FOXQ1: forkhead box Q1
MGGAAGTG	GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2
NNNNNGATANKGNN	GATA1: GATA binding protein 1 (globin transcription factor 1)
CNNRCCCGCATD	GCM1: glial cells missing homolog 1 (Drosophila)
AWAATTRG	HOXA4: homeobox A4
NNNTGAGTCAKCN	JUN: jun oncogene
TGANNYRGCA	NFE2L1: nuclear factor (erythroid-derived 2)-like 1 MAFG: v-maf musculoaponeurotic fibrosarcoma oncogene homolog G (avian)
NTGCTGAGTCAKN	NFE2L2: nuclear factor (erythroid-derived 2)-like 2
TTAYRTAA	NFIL3: nuclear factor, interleukin 3 regulated
CWTAATTG	NKX2-5: NK2 transcription factor related, locus 5 (Drosophila)
NWNAGRACAN	NR3C1: nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
WNTAATCCCAR	PITX2: paired-like homeodomain transcription factor 2
NNGAATATKCANNNN	POU2F1: POU domain, class 2, transcription factor 1
NNNRTAATNANNN	POU2F1: POU domain, class 2, transcription factor 1
TNATTTGCATW	POU2F1: POU domain, class 2, transcription factor 1
NNNNATGCAAATNAN	POU2F1: POU domain, class 2, transcription factor 1
WNNANYYAATTANCNN	PRRX2: paired related homeobox 2
RACCACAR	RUNX1: runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)

NNGKNTGTGGTTWNC	RUNX1: runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)
YGAMCTNNASTRACCYN	RXRB: retinoid X receptor, beta
TGTCTGTCT	SMAD3: SMAD, mothers against DPP homolog 3 (Drosophila)
CCAWATAWGGMNMNG	SRF: serum response factor (c-fos serum response element-binding transcription factor)
NAWTTCYN	STAT5A: signal transducer and activator of transcription 5A
TTCYNRGAA	STAT5B: signal transducer and activator of transcription 5B
WRGTTAATNATTAACNNN	TCF1: transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
NNNNNATGACTCAGCANTTNNG	TCF11 MAFG: v-maf musculoaponeurotic fibrosarcoma oncogene homolog G (avian)
NCACCTGYYNCNKN	TCF3: transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
ATGTTWAYATAA	TEF: thyrotrophic embryonic factor
GATTGGY	Unknown
TTCYRGAA	Unknown
TTTGGGAGR	Unknown
CTNATTTGCATAY	Unknown
TNATTTGCATN	Unknown
TTCCCGKAA	Unknown
NNNNNTTCTKGGA	Unknown

Appendix 14. Transcription factors targeting genes with hypermethylated promoters.

Motif-hypo	Transcription Factor
KDMAYYNTGACCT	AR: androgen receptor (dihydrotestosterone receptor; testicular feminization; spinal and bulbar muscular atrophy; Kennedy disease)
TGAYRTCA	ATF3: activating transcription factor 3
CVTGACGYMABG	ATF4: activating transcription factor 4 (tax-responsive enhancer element B67)
GATTTAACATAA	CDC5L: CDC5 cell division cycle 5-like (S
NNARGNCANNNTGACCYNN	ESR1: estrogen receptor 1
KATTGTTTRTTTW	FOXF2: forkhead box F2
GNNTTGTTTACNTT	FOXO1A: forkhead box O1A (rhabdomyosarcoma)
ANGNDGATAANNGN	GATA1: GATA binding protein 1 (globin transcription factor 1)
WGATARN	GATA1: GATA binding protein 1 (globin transcription factor 1)
NNNGATWANN	GATA6: GATA binding protein 6
SNNNCCNCAGGCN	GTF3A: general transcription factor IIIA
NNNNGNRTCTGGMWTT	HAND1: heart and neural crest derivatives expressed 1
RGAANNTTC	HSF1: heat shock transcription factor 1
AGAANRTTCN	HSF1: heat shock transcription factor 1
CAGTGGG	KLF12: Kruppel-like factor 12
TCAAAG	LEF1: lymphoid enhancer-binding factor 1 TCF1: transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
NMGATANSG	LMO2: LIM domain only 2 (rhombotin-like 1)
CTAWWWATA	MEF2A: MADS box transcription enhancer factor 2, polypeptide A (myocyte enhancer factor 2A)
NNGTTGTTTACNTN	MLLT7: myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 7
TGCCAAR	NF1: neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)
NNTTGGCNNNNNNCCNNN	NF1: neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)
NANWGGAAAANN	NFAT NFATC
TTAAGTRSTT	NKX2-2: NK2 transcription factor related, locus 2 (Drosophila)
RRGGTYANTRNM	NR1H4: nuclear receptor subfamily 1, group H, member 4
NAAWAATTANS	PAX4: paired box gene 4
ANCAATCAW	PBX1: pre-B-cell leukemia transcription factor 1
NNNGGGAGTNNNNS	PCAF: p300/CBP-associated factor
MKVATTTGCATATT	POU2F1: POU domain, class 2, transcription factor 1
RGAGGAARY	SPI1: spleen focus forming virus (SFFV) proviral integration oncogene

	spi1
CACSCCA	SREBF1: sterol regulatory element binding transcription factor 1
SCCAWATAWGGMNMNNNN	SRF: serum response factor (c-fos serum response element-binding transcription factor)
GNCCAWATAWGGMN	SRF: serum response factor (c-fos serum response element-binding transcription factor)
NNNAACAGATGKTNNN	TAL1: T-cell acute lymphocytic leukemia 1 TCF3: transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)
DGTTAATKAWTNACCAM	TCF1: transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor
RCCWGCTG	TCF12: transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)
WGGAATGY	TEAD1: TEA domain family member 1 (SV40 transcriptional enhancer factor)
WGARYCAGCTGYGGNCNK	TFAP4: transcription factor AP-4 (activating enhancer binding protein 4)
RNCAGCTGC	TFAP4: transcription factor AP-4 (activating enhancer binding protein 4)
WCTCAAGTGT	TITF1: thyroid transcription factor 1
NNNNCAAGNRNN	TITF1: thyroid transcription factor 1
NNGTAAKTNG	TLX2: T-cell leukemia homeobox 2
NNNTAATTAGCNNN	VSX1: visual system homeobox 1 homolog, CHX10-like (zebrafish)
NGGGKGGTC	ZIC3: Zic family member 3 heterotaxy 1 (odd-paired homolog, Drosophila)
NNAACATCTGGA	ZNF238: zinc finger protein 238
AAANWWTGC	Unknown
AAAYRNCTG	Unknown
TGGNNNNNNKCCAR	Unknown
RGCTATWTTTAR	Unknown

Appendix 15. KEGG Pathway analysis of genes hypermethylated in cSCC

Description	# of Genes	p-value	Adjusted p value
Pathways in cancer	51	2.85E-17	5.31E-15
Endocytosis	36	7.33E-16	6.82E-14
Regulation of actin cytoskeleton	32	8.13E-11	2.52E-09
Wnt signaling pathway	31	2.15E-14	1.34E-12
Insulin signaling pathway	29	6.41E-14	2.98E-12
MAPK signaling pathway	29	6.45E-07	5.00E-06
Focal adhesion	27	2.02E-08	2.69E-07
Jak-STAT signaling pathway	23	3.42E-08	4.25E-07
Chemokine signaling pathway	21	1.65E-05	8.52E-05
Glycerophospholipid metabolism	20	8.89E-12	3.31E-10
Axon guidance	20	1.26E-07	1.12E-06
Fc gamma R-mediated phagocytosis	19	5.12E-09	8.73E-08
GnRH signaling pathway	19	1.03E-08	1.47E-07
Vascular smooth muscle contraction	19	9.17E-08	8.53E-07
Neurotrophin signaling pathway	19	4.02E-07	3.25E-06
Tight junction	19	1.06E-06	7.27E-06
Calcium signaling pathway	19	6.58E-05	2.99E-04
Cytokine-cytokine receptor interaction	19	8.25E-03	2.10E-02
ErbB signaling pathway	18	5.16E-09	8.73E-08
Chronic myeloid leukemia	17	2.01E-09	4.67E-08
Phosphatidylinositol signaling system	17	3.89E-09	8.04E-08
Prostate cancer	17	4.80E-08	5.58E-07
PPAR signaling pathway	16	6.37E-09	9.88E-08
mTOR signaling pathway	15	7.02E-10	1.87E-08
Pancreatic cancer	15	5.88E-08	6.07E-07
Cell cycle	15	1.33E-04	5.26E-04
Ubiquitin mediated proteolysis	15	3.05E-04	9.78E-04
Acute myeloid leukemia	14	5.15E-08	5.63E-07
Colorectal cancer	14	8.03E-08	7.86E-07
Adipocytokine signaling pathway	14	2.25E-07	1.91E-06
Small cell lung cancer	14	3.96E-06	2.46E-05
Progesterone-mediated oocyte maturation	14	5.27E-06	3.06E-05
TGF-beta signaling pathway	14	5.27E-06	3.06E-05
Oocyte meiosis	14	1.33E-04	5.26E-04
Glioma	13	1.03E-06	7.27E-06
Fc epsilon RI signaling pathway	13	1.00E-05	5.55E-05
Apoptosis	13	3.31E-05	1.60E-04

Melanogenesis	13	1.57E-04	5.83E-04
Toll-like receptor signaling pathway	13	1.57E-04	5.83E-04
Leukocyte transendothelial migration	13	6.60E-04	2.01E-03
Non-small cell lung cancer	12	8.12E-07	6.04E-06
Epithelial cell signaling in Helicobacter pylori infection	12	1.05E-05	5.56E-05
Adherens junction	12	2.93E-05	1.47E-04
VEGF signaling pathway	12	3.35E-05	1.60E-04
Dilated cardiomyopathy	12	2.22E-04	7.66E-04
T cell receptor signaling pathway	12	9.72E-04	2.87E-03
Lysosome	12	2.58E-03	7.28E-03
Purine metabolism	12	2.10E-02	4.70E-02
Glycerolipid metabolism	11	2.09E-06	1.39E-05
Endometrial cancer	11	3.90E-06	2.46E-05
Melanoma	11	8.45E-05	3.60E-04
Gap junction	11	7.04E-04	2.11E-03
Type II diabetes mellitus	10	1.01E-05	5.55E-05
Renal cell carcinoma	10	3.47E-04	1.10E-03
Vasopressin-regulated water reabsorption	9	3.93E-05	1.83E-04
Notch signaling pathway	9	6.80E-05	3.01E-04
Inositol phosphate metabolism	9	2.08E-04	7.31E-04
Basal cell carcinoma	9	2.40E-04	8.13E-04
Hedgehog signaling pathway	9	2.77E-04	9.03E-04
Hypertrophic cardiomyopathy (HCM)	9	5.56E-03	1.50E-02
Hematopoietic cell lineage	9	6.97E-03	1.80E-02
Bladder cancer	8	1.79E-04	6.39E-04
Fatty acid metabolism	8	1.79E-04	6.39E-04
Lysine degradation	8	2.51E-04	8.32E-04
ECM-receptor interaction	8	1.59E-02	3.90E-02
Thyroid cancer	7	9.39E-05	3.88E-04
Vibrio cholerae infection	7	5.64E-03	1.50E-02
NOD-like receptor signaling pathway	7	9.81E-03	2.43E-02
p53 signaling pathway	7	1.71E-02	4.13E-02
Long-term depression	7	1.84E-02	4.25E-02
Long-term potentiation	7	1.84E-02	4.25E-02
Biosynthesis of unsaturated fatty acids	6	1.44E-04	5.60E-04
Glycosaminoglycan biosynthesis - heparan sulfate	6	3.91E-04	1.21E-03
Ether lipid metabolism	6	1.49E-03	4.27E-03
Aldosterone-regulated sodium reabsorption	6	5.26E-03	1.44E-02
Amino sugar and nucleotide sugar metabolism	6	6.63E-03	1.74E-02
Circadian rhythm - mammal	5	8.51E-05	3.60E-04
Glycosaminoglycan biosynthesis - chondroitin sulfate	5	1.30E-03	3.78E-03
Fructose and mannose metabolism	5	9.38E-03	2.36E-02
Pyruvate metabolism	5	1.84E-02	4.25E-02
Aminoacyl-tRNA biosynthesis	5	2.03E-02	4.60E-02

Steroid biosynthesis	4	3.53E-03	9.81E-03
Glycosaminoglycan biosynthesis - keratan sulfate	3	1.85E-02	4.25E-02
Primary bile acid biosynthesis	3	2.21E-02	4.90E-02

Appendix 16. Dysregulated KEGG pathways in genes hypomethylated in cSCC.

Description	# of Genes	p-value	Adjusted p value
Neuroactive ligand-receptor interaction	26	2.36E-12	4.39E-10
Long-term depression	12	2.62E-09	2.37E-07
Pathways in cancer	24	3.82E-09	2.37E-07
Cytokine-cytokine receptor interaction	20	5.10E-08	2.37E-06
Type II diabetes mellitus	9	9.52E-08	3.54E-06
Hypertrophic cardiomyopathy (HCM)	11	2.43E-07	7.54E-06
Focal adhesion	16	4.83E-07	1.28E-05
Dilated cardiomyopathy	10	4.32E-06	9.63E-05
MAPK signaling pathway	17	4.66E-06	9.63E-05
Regulation of actin cytoskeleton	15	5.92E-06	1.08E-04
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	9	6.39E-06	1.08E-04
Melanogenesis	10	1.09E-05	1.69E-04
Calcium signaling pathway	13	1.43E-05	1.94E-04
ECM-receptor interaction	9	1.46E-05	1.94E-04
Endocytosis	13	1.92E-05	2.38E-04
Gap junction	9	2.56E-05	2.97E-04
mTOR signaling pathway	7	2.95E-05	3.23E-04
Hedgehog signaling pathway	7	4.82E-05	4.98E-04
GnRH signaling pathway	9	6.40E-05	6.26E-04
Cell adhesion molecules (CAMs)	10	1.15E-04	1.01E-03
Tight junction	10	1.15E-04	1.01E-03
Chemokine signaling pathway	12	1.23E-04	1.04E-03
Insulin signaling pathway	10	1.38E-04	1.11E-03
Adipocytokine signaling pathway	7	1.54E-04	1.19E-03
Vascular smooth muscle contraction	9	1.74E-04	1.30E-03
Fructose and mannose metabolism	5	2.73E-04	1.95E-03
Neurotrophin signaling pathway	9	3.45E-04	2.38E-03
Axon guidance	9	4.11E-04	2.73E-03
Fc epsilon RI signaling pathway	7	4.29E-04	2.75E-03
Glycolysis / Gluconeogenesis	6	6.87E-04	4.26E-03
ErbB signaling pathway	7	7.68E-04	4.61E-03
Glycosphingolipid biosynthesis - lacto and neolacto series	4	9.59E-04	5.57E-03
Wnt signaling pathway	9	1.27E-03	7.14E-03
Long-term potentiation	6	1.31E-03	7.14E-03
Melanoma	6	1.41E-03	7.47E-03
Histidine metabolism	4	1.46E-03	7.55E-03

Jak-STAT signaling pathway	9	1.52E-03	7.65E-03
Glyoxylate and dicarboxylate metabolism	3	2.40E-03	1.17E-02
Basal cell carcinoma	5	2.54E-03	1.21E-02
Progesterone-mediated oocyte maturation	6	3.72E-03	1.69E-02
TGF-beta signaling pathway	6	3.72E-03	1.69E-02
NOD-like receptor signaling pathway	5	4.28E-03	1.89E-02
Aldosterone-regulated sodium reabsorption	4	5.78E-03	2.50E-02
Fc gamma R-mediated phagocytosis	6	6.68E-03	2.82E-02
Pancreatic cancer	5	7.16E-03	2.89E-02
Renal cell carcinoma	5	7.16E-03	2.89E-02
Toll-like receptor signaling pathway	6	8.47E-03	3.35E-02
Ubiquitin mediated proteolysis	7	1.01E-02	3.92E-02
Endometrial cancer	4	1.22E-02	4.60E-02
Cardiac muscle contraction	5	1.24E-02	4.60E-02
Thyroid cancer	3	1.33E-02	4.84E-02
Inositol phosphate metabolism	4	1.39E-02	4.97E-02

Appendix 17. Differentially methylated probes in cSCC compared to AK.

cgID	dm	Gene
cg01025800	-0.200396164	PTBP1
cg19764731	-0.195589479	0
cg14209920	-0.18358751	NFIX
cg18560551	-0.182056302	EHF
cg07636225	-0.182024189	RTN3
cg25138553	-0.178570093	HSPG2
cg16853770	-0.178547855	VPS37B
cg16324314	-0.177755179	0
cg10126234	-0.176811734	ZFYVE21
cg16209517	-0.175021566	KIAA1737
cg10402698	-0.172310962	SMAD6
cg00302587	-0.171827787	NCOA4
cg20984972	-0.171667827	NFIC
cg12914733	-0.170301963	AP2A2
cg08545463	-0.167905792	EXPH5
cg04646674	-0.167871478	NFIX
cg25580656	-0.165083285	ZFYVE21
cg24153924	-0.162354241	0
cg10333594	-0.160693784	CRAMP1L
cg08983011	-0.157025628	BCL9L
cg24671153	-0.156263785	AGPAT1
cg15926004	-0.15517117	C1orf133; SERTAD4
cg05241828	-0.155010707	GPT
cg08761535	-0.152455719	ZNRF1
cg00510149	-0.151391703	MRVI1
cg06026545	-0.150234961	ACACA
cg07962882	-0.14816068	SBNO2
cg09957864	-0.147409351	GPT
cg21467371	-0.146296511	ARL16
cg04095826	-0.145737354	PGBD5
cg16130802	-0.145673851	BCL9L
cg01500055	-0.14358196	0
cg24182798	-0.142043711	0
cg26995083	-0.139705263	TJAP1
cg12280150	-0.138200533	HEPHL1
cg25312122	-0.138024814	PEA15
cg07259418	-0.137423522	PRICKLE2
cg12150931	-0.137081777	ZNF385A

cg18378955	-0.135948368	0
cg06182584	-0.133887351	C6orf136
cg17611512	-0.132961384	NCRNA00175; COL18A1
cg04984818	-0.132273499	ZFYVE21
cg25724837	-0.13167157	0
cg08648499	-0.13140215	PRSS8
cg02832224	-0.128587843	SSBP3
cg26788852	-0.12835503	MAPKBP1
cg25600446	-0.127606936	GPT
cg02681400	-0.127387518	BRD4
cg25352281	-0.126522736	PPP2R2D
cg07658280	-0.124536636	GPT
cg01771850	-0.123739187	TBCD
cg11634930	-0.123713455	MKNK2
cg17858643	-0.12321625	LIMD1
cg15801751	-0.122893301	MARK2
cg14010829	-0.118459422	ASS1
cg21772826	-0.116155018	TPRG1
cg03474702	-0.116115425	IMPA2
cg14711433	-0.115242424	EXTL3
cg00458395	-0.113438917	0
cg26543333	-0.113179716	NFIC
cg01906801	-0.112316975	MED24
cg13154413	-0.111544619	ZFYVE21
cg05578673	-0.111421553	LRRC8D
cg04842880	-0.109849412	CASZ1
cg13243168	-0.109262749	SMARCD2
cg22940022	-0.10919705	MCM7
cg21241411	-0.108903167	RGL2
cg08727352	-0.10504797	RGL2
cg25050723	-0.104796521	HCCA2
cg04085707	-0.103673315	MAPK8IP3
cg19316405	-0.103147523	PDDC1
cg23875758	-0.10215405	SREBF1
cg21098787	-0.102141888	BRD4
cg13169667	-0.102093105	ZC3H3
cg00545886	-0.10202929	BLCAP
cg03691756	-0.101259196	0
cg19697795	-0.101115726	0
cg04902929	-0.100627197	C1orf203
cg25609878	-0.099414231	0
cg15231749	-0.097980564	RGL2
cg11835619	-0.096356557	ERBB3
cg27575890	-0.09367424	EIF3D
cg17483792	-0.093262717	AHCYL2
cg25384949	-0.092098318	NOTCH4

cg05131707	-0.089124556	PCBP2
cg00033551	-0.084966555	MGRN1
cg06651452	-0.08304793	ZBTB7A
cg01120851	-0.082119043	KLHL21
cg00280345	-0.078973312	GPT
cg07918736	-0.078817878	ILF3
cg04594691	-0.077344043	0
cg01678580	-0.076066838	MGRN1
cg07589968	-0.0750991	KCNS1
cg05709598	-0.073360333	KIAA0182
cg14297023	-0.070897999	EIF3D
cg13219127	-0.06731669	TOLLIP
cg02924834	-0.065951124	KLHDC4
cg01320211	-0.065014444	SEMA6D
cg14152306	-0.064739891	BAG5
cg26881362	-0.061450279	ACADVL
cg00287016	-0.058817462	MCM7
cg01721450	-0.055341182	SEC61A1
cg01366338	-0.055036686	ALDH4A1
cg12218406	-0.053175661	ARRDC2
cg23092788	-0.049532195	TMEM175
cg20166919	-0.049411165	FAM53B
cg02012771	-0.037340442	NEIL1
cg25594549	-0.030273759	MED26
cg12296218	0.003293619	DUT
cg21110939	0.005877681	SV2B
cg13467135	0.015214722	KIAA1598
cg22933133	0.033968768	0
cg26389255	0.034687528	0
cg01192112	0.041782255	PHACTR2
cg09123524	0.045813779	TNXB
cg05344955	0.046286934	0
cg16347868	0.058811498	PBX1
cg25194822	0.068693261	KCNN2
cg25181684	0.078390378	SH3PXD2A
cg05226607	0.081722682	SPINK5L2
cg21297772	0.100578072	SLC5A9
cg02854972	0.102930082	0
cg02388849	0.105753771	0
cg03110996	0.107758314	0
cg16278514	0.108840244	0
cg25446789	0.110247184	DTNB
cg24189745	0.11098271	0
cg18151345	0.111259861	SLC15A3
cg22013055	0.115167741	LOC100134259
cg11749902	0.125677021	0

cg07586285	0.126796102	0
cg04043538	0.127468562	0
cg00251358	0.129096011	0
cg03033182	0.142986067	SCOC
cg23817637	0.147640696	CLRN3
cg23403895	0.148294837	PALLD
cg17841267	0.154152093	0
cg04233620	0.154557754	ADARB1
cg24943066	0.157966566	0
cg20294319	0.168065699	LRCH1
cg20694147	0.169645139	MTHFD2
cg03652336	0.183082351	MAML3
cg10166664	0.188049049	0
cg07869023	0.189498052	PCSK2
cg16346032	0.198072208	0
cg10575547	0.203364511	TRERF1
cg18826637	0.22034533	0
cg10344477	0.254253167	B3GNTL1

Appendix 18. Clinical characteristics of study participants that consented to have their lesions examined by SNP arrays (Chapter 3).

Patient	Age at recruitment	Sex	Immune status	Skin type	UV exposure	Tumours	
1	38	M	Immunosuppressed	II	medium		5 SCC
2	57	F	Immunosuppressed	II	medium	1 BCC	6 SCC
3	58	M	Immunosuppressed	I	medium	6 BCC	
4	64	M	Immunosuppressed	II	medium	5 BCC	29 SCC
5	71	F	Immunosuppressed	II	medium		1 SCC
6	76	F	Immunosuppressed	I	medium	1 BCC	2 SCC
7	57	M	Immunosuppressed	II	medium	3 BCC	4 SCC
8	58	M	Immunosuppressed	I	high	14 BCC	3 SCC
9	70	M	Immunosuppressed	I	medium	23 BCC	29 SCC
10	70	M	Immunosuppressed	I	medium		1 SCC
11	51	M	Immunosuppressed	I	medium		2 SCC
12	73	M	Immunosuppressed	II	high	2 BCC	6 SCC
13	66	M	Immunosuppressed	I	high		1 SCC
14	69	M	Immunosuppressed	I	high		4 SCC
15	79	M	Immunosuppressed	I	high	1 BCC	2 SCC
16	52	F	Immunosuppressed	II	high	1 BCC	6 SCC
17	54	F	Immunosuppressed	I	medium		4 SCC
18	83	M	Immunosuppressed	I	high	8 BCC	9 SCC
19	59	M	Immunosuppressed	II	high		6 SCC
20	64	M	Immunosuppressed	I	medium	1 BCC	2 SCC
21	61	F	Immunosuppressed	II	medium	2 BCC	5 SCC
22	68	F	Immunosuppressed	I	medium	1 BCC	
23	71	F	Immunosuppressed	II	medium	4 BCC	2 SCC
24	80	M	Immunosuppressed	II	high	6 BCC	1 SCC
25	80	M	Immunosuppressed	I	medium	5 BCC	2 SCC
26	45	M	Immunosuppressed	II	medium		6 SCC
27	68	M	Immunosuppressed	I	high	1 BCC	9 SCC
28	47	M	Immunosuppressed	I	medium		3 SCC
29	65	M	Immunosuppressed	II	high		8 SCC
30	58	M	Immunosuppressed	II	medium	5 BCC	11 SCC
31	57	M	Immunosuppressed	II	medium	8 BCC	2 SCC
32	50	M	Immunocompetent	I	Not known	2 BCC	2 SCC
33	85	M	Immunocompetent	I	Not known	3 BCC	
34	84	M	Immunocompetent	I	Not known		1 SCC
35	75	M	Immunocompetent	I	Not known	5 BCC	4 SCC
36	74	M	Immunocompetent	I	Not known	4 BCC	2 SCC
37	80	M	Immunocompetent	I	Not known	4 BCC	2 SCC

38	77	M	Immunocompetent	II	Not known	2 BCC	
39	83	M	Immunocompetent	II	Not known		1 SCC
40	82	M	Immunocompetent	II	Not known		1 SCC
41	77	M	Immunocompetent	I	Not known	1 BCC	
42	89	M	Immunocompetent	I	Not known	3 BCC	1 SCC
43	71	M	Immunocompetent	I	Not known		4 SCC
44	71	M	Immunocompetent	I	Not known	4 BCC	
45	82	M	Immunocompetent	I	Not known		1 SCC
46	85	M	Immunocompetent	I	Not known		2 SCC
47	85	M	Immunocompetent	I	Not known	2 BCC	1 SCC
48	71	M	Immunocompetent	II	Not known		1 SCC
49	80	M	Immunocompetent	I	Not known	1 BCC	1 SCC
50	79	M	Immunocompetent	I	Not known	none	
51	85	F	Immunocompetent	I	Not known	1 BCC	2 SCC
52	52	M	Immunocompetent	I	Not known		2 SCC
53	70	F	Immunocompetent	II	Not known		2 SCC

Appendix 19. Clinical characteristics of samples used for expression microarray validation

Sample	Gender	Site	Age at biopsy	Immune Status
WD1	F	forehead	75	IC
WD2	F	calf	44	RT
WD3	M	leg	58	IC
WD4	M	scalp	49	RT
WD5	M	Forehead	53	RT
WD6	M	ear	63	RT
WD7	M	lower leg	59	RT
WD8	M	calf	49	RT
WD9	F	cheek	51	RT
MD1	M	scalp	64	IC
MD2	M	forehead	66	IC
MD3	M	ear	65	RT
MD4	F	finger	71	RT
MD5	M	ear	76	RT
MD6	F	dorsal hand	64	RT
MD7	M	neck	66	RT
MD8	M	shoulder	61	RT
MD9	M	forearm	71	RT
AK1	M	finger	42	RT
AK2	M	scalp	63	RT
AK3	F	arm	56	RT
AK4	M	chest	63	RT
AK5	F	temple	74	IC
AK6	F	ear	46	RT
AK7	M	dorsal hand	51	RT
AK8	M	arm	62	RT
AK9	M	ear	78	IC
SE1	M	forearm	46	RT
SE2	M	forearm	78	IC
SE3	F	forearm	60	RT
SE4	M	forearm	64	RT
SE5	M	forearm	48	RT
NSE1	M	inner arm	62	RT

NSE2	M	inner arm	63	RT
NSE3	M	buttock	57	RT
NSE4	M	inner arm	79	IC

Appendix 20. Average methylation levels in microdissected paired cSCC and peritumoural skin based on bisulfite sequencing

Gene	Name	Skin	CSCC
ENSG00000173040	EVC2	0.375	0.57381
ENSG00000119698	PPP4R4	0.80258	0.835
ENSG00000157036	EXOG	0.421429	0.414914
ENSG00000102385	DRP2	0.770536	0.568452
ENSG00000181215	C4orf50	0.641303	0.63149
ENSG00000087076	HSD17B14	0.274093	0.262222
ENSG00000163788	SNRK	0.324745	0.0545185
ENSG00000171224	C10orf35	0.483334	0.33811
ENSG00000115386	REG1A	0.805769	0.703297
ENSG00000213215	OR2F1	0.691867	0.589744
ENSG00000182931	WFDC10B	0.85	0.63125
ENSG00000163586	FABP1	0.911905	0.817766
ENSG00000171812	COL8A2	0.733333	0.778439
ENSG00000123453	SARDH	0.633333	0.866327
ENSG00000128309	MPST	0.243056	0.5
ENSG00000123171	CCDC70	0.658605	0.767928
ENSG00000134030	CTIF	0.60625	0.425238
ENSG00000171119	NRTN	0.643147	0.706899
ENSG00000109787	KLF3	0.25	0.25
ENSG00000176236	C10orf111	0.165233	0.190322
ENSG00000075142	SRI	0.947829	0.655555
ENSG00000043093	DCUN1D1	0.923737	0.848672

ENSG00000152910	CNTNAP4	0.38795	0.292509
ENSG00000123374	CDK2	0.223333	0.309524
ENSG00000136099	PCDH8	0.0887445	0.296627
ENSG00000147144	CCDC120	0.8	0.797619
ENSG00000008311	AASS	0.272506	0.21309
ENSG00000120251	GRIA2	0.836863	0.831121
ENSG00000136560	TANK	0.687332	0.541883
ENSG00000053702	NRIP2	0.765625	0.572288
ENSG00000109654	TRIM2	0.749074	0.549327
ENSG00000071626	DAZAP1	0.166667	0.0833333
ENSG00000186298	PPP1CC	0.139267	0.112324
ENSG00000003400	CASP10	0.0845828	0.467477
ENSG00000188191	PRKAR1B	0.5411	0.35988
ENSG00000133466	C1QTNF6	0.4	0.52381
ENSG00000134516	DOCK2	0.783413	0.704545
ENSG00000100285	NEFH	0.299273	0.175947
ENSG00000188015	S100A3	0.553333	0.9
ENSG00000198920	KIAA0753	0.203175	0.215
ENSG00000168329	CX3CR1	0.485853	0.682954
ENSG00000126950	TMEM35	0.451275	0.50492
ENSG00000100884	CPNE6	0.541667	0.56875
ENSG00000100053	CRYBB3	0.583333	0.757143
ENSG00000124827	GCM2	0.599459	0.410016
ENSG00000107758	PPP3CB	0.2163	0.255952
ENSG00000050730	TNIP3	0.651235	0.673136
ENSG00000077782	FGFR1	0.795	0.814286
ENSG00000157211	CDCP2	0.845	0.620122
ENSG00000115129	TP53I3	0.614374	0.698901
ENSG00000169314	C22orf15	0.874224	0.785185

ENSG00000157214	STEAP2	0.872222	0.72735
ENSG00000197928	ZNF677	0.486012	0.32681
ENSG00000166069	TMCO5A	0.835318	0.785714
ENSG00000213339	QTRT1	0.232402	0.444796
ENSG00000066629	EML1	0.508333	0.5
ENSG00000205426	KRT81	0.913753	0.769006
ENSG00000177663	IL17RA	0.629609	0.683333
ENSG00000136895	GARNL3	0.200556	0.316182
ENSG00000136574	GATA4	0.416399	0.263889
ENSG00000100565	C14orf166B	0.749675	0.704365
ENSG00000107742	SPOCK2	0.529142	0.389369
ENSG00000157005	SST	0.51462	0.350777
ENSG00000174740	PABPC5	0.319841	0.462434
ENSG00000160991	ORAI2	0.320833	0.55
ENSG00000100739	BDKRB1	0.560714	0.627088
ENSG00000167978	SRRM2	0.740747	0.781944
ENSG00000167874	TMEM88	0.703965	0.769676
ENSG00000023191	RNH1	0.836667	0.913195
ENSG00000145794	MEGF10	0.344758	0.420737
ENSG00000143032	BARHL2	0.330197	0.343574
ENSG00000166961	MS4A15	0.702966	0.725564
ENSG00000137955	RABGGTB	0.282357	0.182374
ENSG00000181781	ODF3L2	0.460582	0.778616
ENSG00000173597	SULT1B1	0.930289	0.961539
ENSG00000188581	KRTAP1-1	0.68254	0.657143
ENSG00000071054	MAP4K4	0.291667	0.407857
ENSG00000180509	KCNE1	0.661668	0.597583
ENSG00000183778	B3GALT5	0.633734	0.643462
ENSG00000112214	FHL5	0.555671	0.604305

ENSG00000128739	SNRPN	0.627629	0.480176
ENSG00000181374	CCL13	0.933036	0.771429
ENSG00000134760	DSG1	0.383643	0.348786
ENSG00000183640	KRTAP8-1	0.750303	0.669898
ENSG00000137090	DMRT1	0.577579	0.516667
ENSG00000162407	PPAP2B	0.240874	0.164154
ENSG00000119973	PRLHR	0.413046	0.2
ENSG00000128040	SPINK2	0.349213	0.500539
ENSG00000100031	GGT1	0.659524	0.488095
ENSG00000184117	NIPSNAP1	0.140278	0.25
ENSG00000158691	ZSCAN12	0.391272	0.356385
ENSG00000166927	MS4A7	0.513095	0.383041
ENSG00000179407	DNAJB8	0.818182	0.59861
ENSG00000161958	FGF11	0.246354	0.410678
ENSG00000115042	FAHD2A	0.5	0.631429
ENSG00000156345	CDK20	0.664262	0.793931
ENSG00000105671	DDX49	0.0807144	0.457407
ENSG00000135898	GPR55	0.827614	0.860019
ENSG00000108688	CCL7	0.486111	0.522222
ENSG00000110628	SLC22A18	0.721429	0.786436
ENSG00000172995	ARPP21	0.531279	0.396635
ENSG00000145107	TM4SF19	0.466244	0.666943
ENSG00000173757	STAT5B	0.458334	0.225
ENSG00000118707	TGIF2	0.33254	0.461508
ENSG00000169026	MFSD7	0.799713	0.554167
ENSG00000123689	G0S2	0.1	0.221795
ENSG00000198183	BPIFA1	0.782027	0.801136
ENSG00000006071	ABCC8	0.240909	0.426667
ENSG00000172346	CSDC2	0.496032	0.747024

ENSG00000179909	ZNF154	0.533812	0.381667
ENSG00000104537	ANXA13	0.702852	0.756594
ENSG00000189401	OTUD6A	0.896429	0.815385
ENSG00000090920	FCGBP	0.740303	0.777778
ENSG00000102967	DHODH	0.330631	0.58425
ENSG00000143194	MAEL	0.727231	0.716932
ENSG00000130340	SNX9	0.535715	0.258179
ENSG00000152601	MBNL1	0.715737	0.721548
ENSG00000139132	FGD4	0.136607	0.217024
ENSG00000126549	STATH	0.680633	0.711194
ENSG00000137976	DNASE2B	0.876936	0.741816
ENSG00000109061	MYH1	0.674151	0.696355
ENSG00000166948	TGM6	0.841987	0.816239
ENSG00000135480	KRT7	0.630769	0.588095
ENSG00000106648	GALNTL5	0.792087	0.808968
ENSG00000156096	UGT2B4	0.4875	0.531566
ENSG00000142513	ACPT	0.705426	0.875155
ENSG00000155495	MAGEC1	1	1
ENSG00000116016	EPAS1	0.892857	0.817272
ENSG00000204787	REG1P	0.625714	0.907936
ENSG00000143479	DYRK3	0.238533	0.224111
ENSG00000101076	HNF4A	0.667304	0.730914
ENSG00000100191	SLC5A4	0.803128	0.782252
ENSG00000176979	TRIM60	0.887984	0.888205
ENSG00000131730	CKMT2	0.8125	0.645714
ENSG00000185621	LMLN	0.12384	0.202532
ENSG00000101843	PSMD10	0.1	0.237075
ENSG00000184922	FMNL1	0.516667	0.659326
ENSG00000141977	CIB3	0.729167	0.401786

ENSG00000152977	ZIC1	0.19246	0.205506
ENSG00000158186	MRAS	0.389593	0.313331
ENSG00000197681	TBC1D3	0.66157	0.699013
ENSG00000105695	MAG	0.416239	0.415661
ENSG00000124134	KCNS1	0.555622	0.616667
ENSG00000188089	PLA2G4E	0.75696	0.815992
ENSG00000064012	CASP8	0.33349	0.458256
ENSG00000121933	ADORA3	0.606429	0.652619
ENSG00000113100	CDH9	0.841901	0.835137
ENSG00000122691	TWIST1	0.313244	0.0227273
ENSG00000155980	KIF5A	0.694286	0.539935
ENSG00000173714	WFIKKN2	0.788889	0.822222
ENSG00000183674	LINC00518	0.472559	0.662842
ENSG00000115896	PLCL1	0.201284	0.172619
ENSG00000151376	ME3	0.435714	0.398571
ENSG00000124140	SLC12A5	0.269392	0.320933
ENSG00000151789	ZNF385D	0.50453	0.336578
ENSG00000086289	EPDR1	0.680482	0.614013
ENSG00000161572	LYZL6	0.668791	0.731823
ENSG00000204897	KRT25	0.395833	0.81713
ENSG00000100321	SYNGR1	0.238095	0.593687
ENSG00000175229	GAL3ST3	0.9375	0.643333
ENSG00000163207	IVL	0.672421	0.7
ENSG00000065029	ZNF76	0.495059	0.515031
ENSG00000130167	TSPAN16	0.54482	0.574934
ENSG00000101441	CST4	0.416667	0.6625
ENSG00000112769	LAMA4	0.375789	0.340317
ENSG00000136869	TLR4	0.608164	0.620553
ENSG00000117298	ECE1	0.25	0.575

ENSG00000189320	FAM180A	0.95	0.609722
ENSG00000138193	PLCE1	0.533541	0.539285
ENSG00000196136	SERPINA3	0.937231	0.790641
ENSG00000104970	KIR3DX1	0.809042	0.493032
ENSG00000175426	PCSK1	0.263262	0.254901
ENSG00000143867	OSR1	0.418889	0.56
ENSG00000167363	FN3K	0.457143	0.655291
ENSG00000130176	CNN1	0.779762	0.490151
ENSG00000185686	PRAME	0.626389	0.690284
ENSG00000167670	CHAF1A	0.0824074	0.115278
ENSG00000028137	TNFRSF1B	0.50974	0.686963
ENSG00000085265	FCN1	0.444444	0.780556
ENSG00000139914	FITM1	0.636233	0.800463
ENSG00000166527	CLEC4D	0.587446	0.538195
ENSG00000140932	CMTM2	0.80071	0.597043
ENSG00000158555	GDPD5	0.53198	0.526191
ENSG00000163156	SCNM1	0.694394	0.566603
ENSG00000149735	GPHA2	0.851191	0.821429
ENSG00000004534	RBM6	0.616667	0.62963
ENSG00000112659	CUL9	0.503987	0.473642
ENSG00000253159	PCDHGA12	0.473297	0.373547
ENSG00000164508	HIST1H2AA	0.863039	0.852002
ENSG00000166793	YPEL4	0.237922	0.219413
ENSG00000117971	CHRNB4	0.772511	0.843878
ENSG00000119737	GPR75	0.844823	0.642337
ENSG00000101473	ACOT8	0.531061	0.64425
ENSG00000162367	TAL1	0.543403	0.272421
ENSG00000112115	IL17A	0.814168	0.753266
ENSG00000153820	SPHKAP	0.37698	0.252194

ENSG00000121481	RNF2	0.401992	0.573006
ENSG00000152583	SPARCL1	0.542747	0.529038
ENSG00000183304	FAM9A	0.833333	0.714286
ENSG00000129221	AIPL1	0.861111	0.839087
ENSG00000132932	ATP8A2	0.662637	0.663889
ENSG00000112053	SLC26A8	0.589849	0.709762
ENSG00000196344	ADH7	0.397186	0.503788
ENSG00000116199	FAM20B	0.339598	0.430333
ENSG00000145888	GLRA1	0.332209	0.166837
ENSG00000183625	CCR3	0.733889	0.724206
ENSG00000188629	ZNF177	0.450099	0.394159
ENSG00000107560	RAB11FIP2	0.0656249	0.376675
ENSG00000198046	ZNF667	0.5	0.39983
ENSG00000136819	C9orf78	0.165179	0.401786
ENSG00000102547	CAB39L	0.155102	0.430952
ENSG00000179817	MRGPRX4	0.694921	0.750469
ENSG00000116191	RALGPS2	0.751786	0.701026
ENSG00000172057	ORMDL3	0.255952	0.254762
ENSG00000066382	MPPED2	0.659722	0.347283
ENSG00000132405	TBC1D14	0.608333	0.5625
ENSG00000176928	GCNT4	0.821429	0.654239
ENSG00000172350	ABCG4	0.292532	0.308333
ENSG00000181638	ZFP41	0.162101	0.0911172
ENSG00000123104	ITPR2	0.136657	0.229279
ENSG00000147687	TATDN1	0.275788	0.372407
ENSG00000178187	ZNF454	0.57465	0.5
ENSG00000114200	BCHE	0.496004	0.5188
ENSG00000102934	PLLP	0.421474	0.638515
ENSG00000196542	SPTSSB	0.591843	0.761823

ENSG00000125965	GDF5	0.449711	0.504792
ENSG00000112246	SIM1	0.292262	0.225062
ENSG00000166501	PRKCB	0.512346	0.200661
ENSG00000152749	GPR180	0.299287	0.372301
ENSG00000185565	LSAMP	0.399662	0.294836
ENSG00000171954	CYP4F22	0.270269	0.377899
ENSG00000136827	TOR1A	0.462208	0.811984
ENSG00000188613	NANOS1	0.25	0.8
ENSG00000182240	BACE2	0.225397	0.554879
ENSG00000120896	SORBS3	0.779546	0.896258
ENSG00000178695	KCTD12	0.218707	0.1875
ENSG00000135636	DYSF	0.360399	0.236667
ENSG00000075826	SEC31B	0.497619	0.546212
ENSG00000203710	CR1	0.518983	0.509613
ENSG00000188386	PPP3R2	0.761235	0.760268
ENSG00000165795	NDRG2	0.230159	0
ENSG00000112337	SLC17A2	0.781041	0.780062
ENSG00000149311	ATM	0.037037	0.107738
ENSG00000038945	MSR1	0.70527	0.801996
ENSG00000103546	SLC6A2	0.563136	0.725866
ENSG00000128285	MCHR1	0.665	0.841856
ENSG00000162972	C2orf47	0.379018	0.0971381
ENSG00000239169	SNORD109B	0.572421	0.561868
ENSG00000124574	ABCC10	0.675	0.430556
ENSG00000185883	ATP6V0C	0.7	0.75
ENSG00000028277	POU2F2	0.693734	0.721889
ENSG00000073754	CD5L	0.563581	0.769994
ENSG00000108556	CHRNE	0.382738	0.597959
ENSG00000205358	MT1H	0.8125	0.764253

ENSG00000152672	CLEC4F	0.893333	0.809709
ENSG00000164761	TNFRSF11B	0.277748	0.327521
ENSG00000253953	PCDHGB4	0.45811	0.354898
ENSG00000137674	MMP20	0.828691	0.700219
ENSG00000187323	DCC	0.205436	0.249017
ENSG00000116478	HDAC1	0.373486	0.49079
ENSG00000125888	BANF2	0.609619	0.558677
ENSG00000180138	CSNK1A1L	0.875794	0.86009
ENSG00000182631	RXFP3	0.521516	0.458081
ENSG00000173535	TNFRSF10C	0.8	0.678449
ENSG00000146834	MEPCE	0.316942	0.3916
ENSG00000146731	CCT6A	0.094697	0.26433
ENSG00000033627	ATP6V0A1	0.430296	0.448106
ENSG00000182983	ZNF662	0.827679	0.453897
ENSG00000149435	GGTLC1	0.959375	0.833
ENSG00000127603	MACF1	0.571925	0.754945
ENSG00000180035	ZNF48	0.255626	0.332565
ENSG00000136114	THSD1	0.688624	0.693753
ENSG00000103569	AQP9	0.392744	0.500641
ENSG00000125492	BARHL1	0.24854	0.198046
ENSG00000264424	MYH4	0.671627	0.733929
ENSG00000134256	CD101	0.823418	0.751754
ENSG00000106331	PAX4	0.8125	0.451191
ENSG00000173264	GPR137	0.25	0.25
ENSG00000022267	FHL1	0.2	0.166667
ENSG00000064601	CTSA	0.660715	0.660715
ENSG00000177138	FAM9B	0.808333	0.666664
ENSG00000104883	PEX11G	0.877273	0.629941
ENSG00000186970	KRTAP15-1	0.763134	0.662382

ENSG00000188818	ZDHHC11	0.647619	0.469231
ENSG00000170703	TTLL6	0.333331	0.382916
ENSG00000134802	SLC43A3	0.388548	0.357576
ENSG00000072163	LIMS2	0.802745	0.801332
ENSG00000136542	GALNT5	0.49911	0.404896
ENSG00000206190	ATP10A	0.85303	0.875
ENSG00000145945	FAM50B	0.62	0.599921
ENSG00000188001	TPRG1	0.416667	0.260417
ENSG00000172016	REG3A	0.908334	0.62627
ENSG00000136908	DPM2	0.448462	0.525
ENSG00000136546	SCN7A	0.312843	0.272799
ENSG00000154277	UCHL1	0.703635	0.544218
ENSG00000086730	LAT2	0.698512	0.753363
ENSG00000152795	HNRNPDL	0.222222	0.133333
ENSG00000169340	PDILT	0.758841	0.736433
ENSG00000099330	OCEL1	0.342108	0.550323
ENSG00000177143	CETN1	0.802271	0.85114
ENSG00000186965	KRTAP19-2	0.971282	0.692505
ENSG00000171405	XAGE5	0.808333	0.76712
ENSG00000127533	F2RL3	0.885714	0.65
ENSG00000254122	PCDHGB7	0.542009	0.45617
ENSG00000168917	SLC35G2	0.888889	0.612127
ENSG00000115919	KYNU	0.698956	0.45592
ENSG00000117016	RIMS3	0.479762	0.36
ENSG00000059804	SLC2A3	0.444288	0.453846
ENSG00000091844	RGS17	0.619048	0.42
ENSG00000180152	AC079753.4	0.690477	0.764812
ENSG00000269699	ZIM2	0.560328	0.568783
ENSG00000149090	PAMR1	0.604858	0.520924

ENSG00000108700	CCL8	0.712743	0.783593
ENSG00000103056	SMPD3	0.679412	0.816667
ENSG00000179676	LINC00305	0.853401	0.736667
ENSG00000186564	FOXD2	0.575635	0.473545
ENSG00000146197	SCUBE3	0.497561	0.329236
ENSG00000012048	BRCA1	0.0982249	0.499335
ENSG00000127529	OR7C2	0.759127	0.762047
ENSG00000167749	KLK4	0.756085	0.426389
ENSG00000163879	DNALI1	0.699492	0.614105
ENSG00000177324	BEND2	0.816749	0.780655
ENSG00000154556	SORBS2	0.579071	0.664394
ENSG00000197361	FBXL22	0.752628	0.453241
ENSG00000147166	ITGB1BP2	0.833333	1
ENSG00000160339	FCN2	0.714286	0.833333
ENSG00000164532	TBX20	0.5625	0.272129

Bibliography

ABATE-SHEN, C. 2002. Deregulated homeobox gene expression in cancer: cause or consequence? Nature reviews. Cancer, 2, **,** *777-85.*

ABOUZEID, H., SCHORDERET, D. F., BALMER, A. & MUNIER, F. L. 2009. Germline mutations in retinoma patients: relevance to low-penetrance and low-expressivity molecular basis. Mol Vis, 15, *771-7.*

AHMED, A. A., VIAS, M., IYER, N. G., CALDAS, C. & BRENTON, J. D. 2004. Microarray segmentation methods significantly influence data precision. Nucleic Acids Res, 32, *e50.*

ALAM, M. & RATNER, D. 2001. Cutaneous squamous-cell carcinoma. N Engl J Med, 344, *975-83.*

ALBORES-SAAVEDRA, J., BATICH, K., CHABLE-MONTERO, F., SAGY, N., SCHWARTZ, A. M. & HENSON, D. E. 2010a. Merkel cell carcinoma demographics, morphology, and survival based on 3870 cases: a population based study. Journal of Cutaneous Pathology, 37, *20-27.*

ALBORES-SAAVEDRA, J., BATICH, K., CHABLE-MONTERO, F., SAGY, N., SCHWARTZ, A. M. & HENSON, D. E. 2010b. Merkel cell carcinoma demographics, morphology, and survival based on 3870 cases: a population based study. Journal of cutaneous pathology, 37, *20-7.*

ALONSO, L. & FUCHS, E. 2003. Stem cells of the skin epithelium. Proceedings of the National Academy of Sciences of the United States of America, 100, *11830-11835.*

AMIT, M., CARPENTER, M. K., INOKUMA, M. S., CHIU, C. P., HARRIS, C. P., WAKNITZ, M. A., ITSKOVITZ-ELDOR, J. & THOMSON, J. A. 2000. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. Dev Biol, 227, *271-8.*

ANDERSEN, B., WEINBERG, W. C., RENNEKAMPFF, O., MCEVILLY, R. J., BERMINGHAM, J. R., JR., HOOSHMAND, F., VASILYEV, V., HANSBROUGH, J. F., PITTELKOW, M. R., YUSPA, S. H. & ROSENFELD, M. G. 1997. Functions of the POU domain genes Skn-1a/i and Tst-1/Oct-6/SCIP in epidermal differentiation. Genes Dev, 11, *1873-84.*

ANDERSON, K., LUTZ, C., VAN DELFT, F. W., BATEMAN, C. M., GUO, Y., COLMAN, S. M., KEMPSKI, H., MOORMAN, A. V., TITLEY, I., SWANSBURY, J., KEARNEY, L., ENVER, T. & GREAVES, M. 2011. Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature, 469, *356-61.*

ANDRES, C., BELLONI, B., PUCHTA, U., SANDER, C. A. & FLAIG, M. J. 2010. Prevalence of MCPyV in Merkel cell carcinoma and non-MCC tumors. Journal of Cutaneous Pathology, 37, *28-34.*

ANTEQUERA, F. 2003. Structure, function and evolution of CpG island promoters. Cell Mol Life Sci, 60, *1647-58.*

ANTONSSON, A., WATERBOER, T., BOUWES BAVINCK, J. N., ABENI, D., DE KONING, M., EUVRARD, S., FELTKAMP, M. C. W., GREEN, A. C., HARWOOD, C. A., NALDI, L., NINDL, I., PFISTER, H. J., PROBY, C. M., QUINT, W. G., STOCKFLETH, E., WEISSENBORN, S. J., PAWLITA, M. & NEALE, R. E. 2013. Longitudinal study of seroprevalence and serostability of 34 human papillomavirus types in European organ transplant recipients. Virology, 436, *91-99.*

ANTONSSON, B. E., AVRAMIS, V. I., NYCE, J. & HOLCENBERG, J. S. 1987. Effect of 5-Azacytidine and Congeners on DNA Methylation and Expression of Deoxycytidine Kinase in the

Human Lymphoid Cell Lines CCRF/CEM/0 and CCRF/CEM/dCk. Cancer Research, 47, **,** *3672-3678.*

ARBISER, J. L., FAN, C. Y., SU, X., VAN EMBURGH, B. O., CERIMELE, F., MILLER, M. S., HARVELL, J. & MARINKOVICH, M. P. 2004. Involvement of p53 and p16 tumor suppressor genes in recessive dystrophic epidermolysis bullosa-associated squamous cell carcinoma. J Invest Dermatol, 123, 788-90.

ARORA, A. & SCHOLAR, E. M. 2005. Role of Tyrosine Kinase Inhibitors in Cancer Therapy. Journal of Pharmacology and Experimental Therapeutics, 315, 971-979.

ARRON, S. T., RUBY, J. G., DYBBRO, E., GANEM, D. & DERISI, J. L. 2011. Transcriptome sequencing demonstrates that human papillomavirus is not active in cutaneous squamous cell carcinoma. J Invest Dermatol, 131, 1745-53.

ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A. P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics, 30, 1363-9.

ASHTON, K. J., CARLESS, M. A. & GRIFFITHS, L. R. 2005. Cytogenetic alterations in nonmelanoma skin cancer: a review. Genes Chromosomes Cancer, 43, 239-48.

AUBRY, F. & MACGIBBON, B. 1985. Risk factors of squamous cell carcinoma of the skin. A case-control study in the Montreal region. Cancer, 55, 907-11.

BALASUBRAMANIAN, S., ADHIKARY, G. & ECKERT, R. L. 2010. The Bmi-1 polycomb protein antagonizes the (-)-epigallocatechin-3-gallate-dependent suppression of skin cancer cell survival. Carcinogenesis, 31, 496-503.

BALLESTAR, E. & WOLFFE, A. P. 2001. Methyl-CpG-binding proteins. Targeting specific gene repression. European journal of biochemistry / FEBS, 268, 1-6.

BARBIERI, C. E., TANG, L. J., BROWN, K. A. & PIETENPOL, J. A. 2006. Loss of p63 leads to increased cell migration and up-regulation of genes involved in invasion and metastasis. Cancer Res, 66, 7589-97.

BARNES, C. J., BAGHERI-YARMAND, R., MANDAL, M., YANG, Z., CLAYMAN, G. L., HONG, W. K. & KUMAR, R. 2003. Suppression of epidermal growth factor receptor, mitogen-activated protein kinase, and Pak1 pathways and invasiveness of human cutaneous squamous cancer cells by the tyrosine kinase inhibitor ZD1839 (Iressa). Mol Cancer Ther, 2, 345-51.

BASEMAN, J. G. & KOUTSKY, L. A. 2005. The epidemiology of human papillomavirus infections. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology, 32, 16-24.

BASU, M. & ROY, S. S. 2013. Wnt/beta-catenin pathway is regulated by PITX2 homeodomain protein and thus contributes to the proliferation of human ovarian adenocarcinoma cell, SKOV-3. J Biol Chem, 288, 4355-67.

BAVINCK, J. N. B., NEALE, R. E., ABENI, D., EUVRARD, S., GREEN, A. C., HARWOOD, C. A., DE KONING, M. N. C., NALDI, L., NINDL, I., PAWLITA, M., PFISTER, H., PROBY, C. M., QUINT, W. G. V., TER SCHEGGET, J., WATERBOER, T., WEISSENBORN, S. & FELTKAMP, M. C. W. 2010. Multicenter Study of the Association between Betapapillomavirus Infection and Cutaneous Squamous Cell Carcinoma. Cancer Research, 70, 9777-9786.

BAXTER, R. M. & BRISSETTE, J. L. 2002. Role of the nude gene in epithelial terminal differentiation. J Invest Dermatol, 118, 303-9.

BAYANI, J., BRENTON, J. D., MACGREGOR, P. F., BEHESHTI, B., ALBERT, M., NALLAINATHAN, D., KARASKOVA, J., ROSEN, B., MURPHY, J., LAFRAMBOISE, S., ZANKE, B. & SQUIRE, J. A.

2002. Parallel Analysis of Sporadic Primary Ovarian Carcinomas by Spectral Karyotyping, Comparative Genomic Hybridization, and Expression Microarrays. Cancer Research, 62, ***,*** *3466-3476.*

*BECK, I. M., MULLER, M., MENTLEIN, R., SADOWSKI, T., MUELLER, M. S., PAUS, R. & SEDLACEK, R. 2007. Matrix metalloproteinase-19 expression in keratinocytes is repressed by transcription factors Tst-1 and Skn-1a: implications for keratinocyte differentiation. J Invest Dermatol, 127***,** *1107-14.*

*BECKER, J. C., HOUBEN, R., UGUREL, S., TREFZER, U., PFOHLER, C. & SCHRAMA, D. 2009. MC polyomavirus is frequently present in Merkel cell carcinoma of European patients. The Journal of investigative dermatology, 129***,** *248-50.*

*BELL, C. E., LARIVIÈRE, N. M. K., WATSON, P. H. & WATSON, A. J. 2009. Mitogen-activated protein kinase (MAPK) pathways mediate embryonic responses to culture medium osmolarity by regulating Aquaporin 3 and 9 expression and localization, as well as embryonic apoptosis. Human Reproduction, 24***,** *1373-1386.*

*BEN-DAVID, U. & BENVENISTY, N. 2011. The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat Rev Cancer, 11***,** *268-77.*

*BEN-PORATH, I., THOMSON, M. W., CAREY, V. J., GE, R., BELL, G. W., REGEV, A. & WEINBERG, R. A. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet, 40***,** *499-507.*

*BENJAMIN, C. L. & ANANTHASWAMY, H. N. 2007. p53 and the pathogenesis of skin cancer. Toxicology and Applied Pharmacology, 224***,** *241-248.*

*BERG, R. D. 1996. The indigenous gastrointestinal microflora. Trends Microbiol, 4***,** *430-5.*

*BERMAN, B. & COCKERELL, C. J. 2013. Pathobiology of actinic keratosis: Ultraviolet-dependent keratinocyte proliferation. Journal of the American Academy of Dermatology, 68***,** *S10-S19.*

*BERNSTEIN, B. E., MIKKELSEN, T. S., XIE, X., KAMAL, M., HUEBERT, D. J., CUFF, J., FRY, B., MEISSNER, A., WERNIG, M., PLATH, K., JAENISCH, R., WAGSCHAL, A., FEIL, R., SCHREIBER, S. L. & LANDER, E. S. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell, 125***,** *315-26.*

*BHATIA, K., GOEDERT, J. J., MODALI, R., PREISS, L. & AYERS, L. W. 2010. Immunological detection of viral large T antigen identifies a subset of Merkel cell carcinoma tumors with higher viral abundance and better clinical outcome. International journal of cancer. Journal international du cancer, 127***,** *1493-6.*

BIERHOFF, H., POSTEPSKA-IGIELSKA, A. & GRUMMT, I. 2013. Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics, 9.

*BIGNELL, G. R., HUANG, J., GRESHOCK, J., WATT, S., BUTLER, A., WEST, S., GRIGOROVA, M., JONES, K. W., WEI, W., STRATTON, M. R., FUTREAL, P. A., WEBER, B., SHAPERO, M. H. & WOOSTER, R. 2004. High-Resolution Analysis of DNA Copy Number Using Oligonucleotide Microarrays. Genome Research, 14***,** *287-295.*

*BILBAN, M., BUEHLER, L. K., HEAD, S., DESOYE, G. & QUARANTA, V. 2002. Normalizing DNA microarray data. Curr Issues Mol Biol, 4***,** *57-64.*

*BLANPAIN, C. & FUCHS, E. 2009. Epidermal homeostasis: a balancing act of stem cells in the skin. Nat Rev Mol Cell Biol, 10***,** *207-17.*

*BLANPAIN, C., LOWRY, W. E., PASOLLI, H. A. & FUCHS, E. 2006. Canonical notch signaling functions as a commitment switch in the epidermal lineage. Genes Dev, 20***,** *3022-35.*

BLUEMN, E. G., PAULSON, K. G., HIGGINS, E. E., SUN, Y., NGHIEM, P. & NELSON, P. S. 2009. Merkel cell polyomavirus is not detected in prostate cancers, surrounding stroma, or benign prostate controls. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology, 44, ***,*** *164-6.*

*BOEHNKE, K., FALKOWSKA-HANSEN, B., STARK, H. J. & BOUKAMP, P. 2012. Stem cells of the human epidermis and their niche: composition and function in epidermal regeneration and carcinogenesis. Carcinogenesis, 33****,*** *1247-58.*

*BOHELER, K. R. 2009. Stem cell pluripotency: a cellular trait that depends on transcription factors, chromatin state and a checkpoint deficient cell cycle. J Cell Physiol, 221****,*** *10-7.*

*BOLSTAD, B. M., IRIZARRY, R. A., ASTRAND, M. & SPEED, T. P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19****,*** *185-93.*

*BOTCHEVA, K., MCCORKLE, S. R., MCCOMBIE, W. R., DUNN, J. J. & ANDERSON, C. W. 2011. Distinct p53 genomic binding patterns in normal and cancer-derived human cells. Cell Cycle, 10****,*** *4237-49.*

*BOUMAHDI, S., DRIESSENS, G., LAPOUGE, G., RORIVE, S., NASSAR, D., LE MERCIER, M., DELATTE, B., CAAUWE, A., LENGLEZ, S., NKUSI, E., BROHEE, S., SALMON, I., DUBOIS, C., MARMOL, V. D., FUKS, F., BECK, B. & BLANPAIN, C. 2014. SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. Nature, 511****,*** *246-250.*

*BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 122****,*** *947-56.*

*BOYER, L. A., PLATH, K., ZEITLINGER, J., BRAMBRINK, T., MEDEIROS, L. A., LEE, T. I., LEVINE, S. S., WERNIG, M., TAJONAR, A., RAY, M. K., BELL, G. W., OTTE, A. P., VIDAL, M., GIFFORD, D. K., YOUNG, R. A. & JAENISCH, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature, 441****,*** *349-353.*

*BRAATHEN, L. R., SZEIMIES, R. M., BASSET-SEGUIN, N., BISSONNETTE, R., FOLEY, P., PARISER, D., ROELANDTS, R., WENNBERG, A. M. & MORTON, C. A. 2007. Guidelines on the use of photodynamic therapy for nonmelanoma skin cancer: an international consensus. International Society for Photodynamic Therapy in Dermatology, 2005. J Am Acad Dermatol, 56****,*** *125-43.*

*BRAGADO, P., ESTRADA, Y., PARIKH, F., KRAUSE, S., CAPOBIANCO, C., FARINA, H. G., SCHEWE, D. M. & AGUIRRE-GHISO, J. A. 2013. TGF-beta2 dictates disseminated tumour cell fate in target organs through TGF-beta-RIII and p38alpha/beta signalling. Nat Cell Biol, 15****,*** *1351-61.*

*BRANTSCH, K. D., MEISNER, C., SCHONFISCH, B., TRILLING, B., WEHNER-CAROLI, J., ROCKEN, M. & BREUNINGER, H. 2008. Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study. Lancet Oncol, 9****,*** *713-20.*

*BROWN, P. O. & BOTSTEIN, D. 1999. Exploring the new world of the genome with DNA microarrays. Nat Genet, 21****,*** *33-7.*

*BROWN, V. L., HARWOOD, C. A., CROOK, T., CRONIN, J. G., KELSELL, D. P. & PROBY, C. M. 2004. p16INK4a and p14ARF tumor suppressor genes are commonly inactivated in cutaneous squamous cell carcinoma. J Invest Dermatol, 122****,*** *1284-92.*

*BRUEGGER, C., KEMPF, W., SPOERRI, I., ARNOLD, A. W., ITIN, P. H. & BURGER, B. 2013. MicroRNA expression differs in cutaneous squamous cell carcinomas and healthy skin of immunocompetent individuals. Exp Dermatol, 22***,** *426-8.*

*BUDINSKA, E., POPOVICI, V., TEJPAR, S., D'ARIO, G., LAPIQUE, N., SIKORA, K. O., DI NARZO, A. F., YAN, P., HODGSON, J. G., WEINRICH, S., BOSMAN, F., ROTH, A. & DELORENZI, M. 2013. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. The Journal of Pathology, 231***,** *63-76.*

*BUELL, J. F., TROFE, J., HANAWAY, M. J., BEEBE, T. M., GROSS, T. G., ALLOWAY, R. R., FIRST, M. R. & WOODLE, E. S. 2002. Immunosuppression and Merkel cell cancer. Transplantation proceedings, 34***,** *1780-1.*

*BURGESS, R. J. & ZHANG, Z. 2013. Histone chaperones in nucleosome assembly and human disease. Nat Struct Mol Biol, 20***,** *14-22.*

*BURKHART, D. L. & SAGE, J. 2008. Cellular mechanisms of tumour suppression by the retinoblastoma gene. Nat Rev Cancer, 8***,** *671-682.*

*BURNWORTH, B., ARENDT, S., MUFFLER, S., STEINKRAUS, V., BRÖCKER, E. B., BIREK, C., HARTSCHUH, W., JAUCH, A. & BOUKAMP, P. 2007. The multi-step process of human skin carcinogenesis: A role for p53, cyclin D1, hTERT, p16, and TSP-1. European Journal of Cell Biology, 86***,** *763-780.*

*BUTEL, J. S. 2000. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. Carcinogenesis, 21***,** *405-426.*

*CABRAL, A., FISCHER, D. F., VERMEIJ, W. P. & BACKENDORF, C. 2003. Distinct functional interactions of human Skn-1 isoforms with Ese-1 during keratinocyte terminal differentiation. J Biol Chem, 278***,** *17792-9.*

*CAHILL, N., BERGH, A. C., KANDURI, M., GORANSSON-KULTIMA, H., MANSOURI, L., ISAKSSON, A., RYAN, F., SMEDBY, K. E., JULIUSSON, G., SUNDSTROM, C., ROSEN, A. & ROSENQUIST, R. 2013. 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. Leukemia, 27***,** *150-8.*

*CALIN, G. A., SEVIGNANI, C., DUMITRU, C. D., HYSLOP, T., NOCH, E., YENDAMURI, S., SHIMIZU, M., RATTAN, S., BULLRICH, F., NEGRINI, M. & CROCE, C. M. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A, 101***,** *2999-3004.*

*CALMON, M. F., RODRIGUES, R. V., KANETO, C. M., MOURA, R. P., SILVA, S. D., MOTA, L. D., PINHEIRO, D. G., TORRES, C., DE CARVALHO, A. F., CURY, P. M., NUNES, F. D., NISHIMOTO, I. N., SOARES, F. A., DA SILVA, A. M., KOWALSKI, L. P., BRENTANI, H., ZANELLI, C. F., SILVA, W. A., JR., RAHAL, P., TAJARA, E. H., CARRARO, D. M., CAMARGO, A. A. & VALENTINI, S. R. 2009. Epigenetic silencing of CRABP2 and MX1 in head and neck tumors. Neoplasia, 11***,** *1329-39.*

*CAMPBELL, K. H., MCWHIR, J., RITCHIE, W. A. & WILMUT, I. 1996. Sheep cloned by nuclear transfer from a cultured cell line. Nature, 380***,** *64-6.*

*CAMPBELL, P. J., PLEASANCE, E. D., STEPHENS, P. J., DICKS, E., RANCE, R., GOODHEAD, I., FOLLOWS, G. A., GREEN, A. R., FUTREAL, P. A. & STRATTON, M. R. 2008. Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. Proceedings of the National Academy of Sciences of the United States of America, 105***,** *13081-6.*

CAMPIONE, M., STEINBEISSER, H., SCHWEICKERT, A., DEISSLER, K., VAN BEBBER, F., LOWE, L. A., NOWOTSCHIN, S., VIEBAHN, C., HAFFTER, P., KUEHN, M. R. & BLUM, M. 1999. The *homeobox gene Pitx2: mediator of asymmetric left-right signaling in vertebrate heart and gut looping. Development, 126***,** *1225-1234.*

*CAPASSO, L. L. 2005. Antiquity of cancer. International Journal of Cancer, 113***,** *2-13.*

*CAPECCHI, M. R. 1989. Altering the genome by homologous recombination. Science, 244***,** *1288-92.*

*CARLESS, M. A. & GRIFFITHS, L. R. 2008. Cytogenetics of melanoma and nonmelanoma skin cancer. Adv Exp Med Biol, 624***,** *227-40.*

*CARUANA, G. 2002. Genetic studies define MAGUK proteins as regulators of epithelial cell polarity. Int J Dev Biol, 46***,** *511-8.*

CARVALHO, S., RAPOSO, A. C., MARTINS, F. B., GROSSO, A. R., SRIDHARA, S. C., RINO, J., CARMO-FONSECA, M. & DE ALMEIDA, S. F. 2013. Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Research.

*CASABONNE, D., MICHAEL, K. M., WATERBOER, T., PAWLITA, M., FORSLUND, O., BURK, R. D., TRAVIS, R. C., KEY, T. J. & NEWTON, R. 2007. A prospective pilot study of antibodies against human papillomaviruses and cutaneous squamous cell carcinoma nested in the Oxford component of the European Prospective Investigation into Cancer and Nutrition. International Journal of Cancer, 121***,** *1862-1868.*

*CHANG, A. H. & PARSONNET, J. 2010. Role of Bacteria in Oncogenesis. Clinical Microbiology Reviews, 23***,** *837-857.*

*CHANG, H., QI, X., TRIEU, Y., XU, W., READER, J. C., NING, Y. & REECE, D. 2006. Multiple myeloma patients with CKS1B gene amplification have a shorter progression-free survival post-autologous stem cell transplantation. British Journal of Haematology, 135***,** *486-491.*

*CHATTERJEE, I., HUMTSOE, J. O., KOHLER, E. E., SORIO, C. & WARY, K. K. 2011. Lipid phosphate phosphatase-3 regulates tumor growth via beta-catenin and CYCLIN-D1 signaling. Mol Cancer, 10***,** *51.*

*CHEN, C., WANG, L., LIAO, Q., HUANG, Y., YE, H., CHEN, F., XU, L., YE, M. & DUAN, S. 2013. Hypermethylation of EDNRB promoter contributes to the risk of colorectal cancer. Diagn Pathol, 8***,** *199.*

*CHEN, K. C., WANG, Y. S., HU, C. Y., CHANG, W. C., LIAO, Y. C., DAI, C. Y. & JUO, S. H. 2011. OxLDL up-regulates microRNA-29b, leading to epigenetic modifications of MMP-2/MMP-9 genes: a novel mechanism for cardiovascular diseases. FASEB J, 25***,** *1718-28.*

*CHEN, M. J., YOKOMIZO, T., ZEIGLER, B. M., DZIERZAK, E. & SPECK, N. A. 2009. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. Nature, 457***,** *887-91.*

*CHERPELIS, B. S., MARCUSEN, C. & LANG, P. G. 2002. Prognostic Factors for Metastasis in Squamous Cell Carcinoma of the Skin. Dermatologic Surgery, 28***,** *268.*

*CHEUNG, P. & LAU, P. 2005. Epigenetic regulation by histone methylation and histone variants. Mol Endocrinol, 19***,** *563-73.*

*CHIKH, A., MATIN, R. N. H., SENATORE, V., HUFBAUER, M., LAVERY, D., RAIMONDI, C., OSTANO, P., MELLO-GRAND, M., GHIMENTI, C., BAHTA, A., KHALAF, S., AKGUL, B., BRAUN, K. M., CHIORINO, G., PHILPOTT, M. P., HARWOOD, C. A. & BERGAMASCHI, D. 2011. iASPP/p63 autoregulatory feedback loop is required for the homeostasis of stratified epithelia. EMBO J, 30***,** *4261-4273.*

*CHILES, M. C., AI, L., ZUO, C., FAN, C.-Y. & SMOLLER, B. R. 2003a. E-Cadherin Promoter Hypermethylation in Preneoplastic and Neoplastic Skin Lesions. Mod Pathol, 16***,** *1014-1018.*

*CHILES, M. C., AI, L., ZUO, C., FAN, C. Y. & SMOLLER, B. R. 2003b. E-cadherin promoter hypermethylation in preneoplastic and neoplastic skin lesions. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 16***,** *1014-8.*

*CHOU, B. K., MALI, P., HUANG, X., YE, Z., DOWEY, S. N., RESAR, L. M., ZOU, C., ZHANG, Y. A., TONG, J. & CHENG, L. 2011. Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. Cell Res, 21***,** *518-29.*

*CHOULIARAS, L., MASTROENI, D., DELVAUX, E., GROVER, A., KENIS, G., HOF, P. R., STEINBUSCH, H. W. M., COLEMAN, P. D., RUTTEN, B. P. F. & VAN DEN HOVE, D. L. A. 2013. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. Neurobiology of Aging, 34***,** *2091-2099.*

*CHOWDARY, D., LATHROP, J., SKELTON, J., CURTIN, K., BRIGGS, T., ZHANG, Y., YU, J., WANG, Y. & MAZUMDER, A. 2006. Prognostic gene expression signatures can be measured in tissues collected in RNAlater preservative. J Mol Diagn, 8***,** *31-9.*

*CHU, S. H., FENG, D. F., MA, Y. B., ZHANG, H., ZHU, Z. A., LI, Z. Q. & JIANG, P. C. 2011. Promoter methylation and downregulation of SLC22A18 are associated with the development and progression of human glioma. J Transl Med, 9***,** *156.*

*CIPAK, A., BOROVIC, S., JAGANJAC, M., BRESGEN, N., KIRAC, I., GRBESA, I., MRAKOVCIC, L., CINDRIC, M., SCUKANEC-SPOLJAR, M., GALL-TROSELJ, K., CORIC, M., ECKL, P. & ZARKOVIC, N. 2010. Influence of 4-hydroxynonenal and spleen cells on primary hepatocyte culture and a novel liver-derived cell line resembling hepatocyte stem cells. Acta Biochim Pol, 57***,** *185-91.*

*CLAERHOUT, S., VERSCHOOTEN, L., VAN KELST, S., DE VOS, R., PROBY, C., AGOSTINIS, P. & GARMYN, M. 2010. Concomitant inhibition of AKT and autophagy is required for efficient cisplatin-induced apoptosis of metastatic skin carcinoma. Int J Cancer, 127***,** *2790-803.*

*CLARK, C., PALTA, P., JOYCE, C. J., SCOTT, C., GRUNDBERG, E., DELOUKAS, P., PALOTIE, A. & COFFEY, A. J. 2012. A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip((R)) for methylome profiling. PLoS One, 7***,** *e50233.*

*CLAUDY, A. L., CHIGNOL, M. C. & CHARDONNET, Y. 1989. Detection of herpes simplex virus DNA in a cutaneous squamous cell carcinoma by in situ hybridization. Arch Dermatol Res, 281***,** *333-5.*

*CLAYMAN, G. L., LEE, J. J., HOLSINGER, F. C., ZHOU, X., DUVIC, M., EL-NAGGAR, A. K., PRIETO, V. G., ALTAMIRANO, E., TUCKER, S. L., STROM, S. S., KRIPKE, M. L. & LIPPMAN, S. M. 2005. Mortality Risk From Squamous Cell Skin Cancer. Journal of Clinical Oncology, 23***,** *759-765.*

*CLAYSON, D. B. 1981. Specific aromatic amines as occupational bladder carcinogens. Natl Cancer Inst Monogr***,** *15-9.*

*CLOOS, P. A., CHRISTENSEN, J., AGGER, K. & HELIN, K. 2008. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes Dev, 22***,** *1115-40.*

*COFFER, P., DE JONGE, M., METTOUCHI, A., BINETRUY, B., GHYSDAEL, J. & KRUIJER, W. 1994. junB promoter regulation: Ras mediated transactivation by c-Ets-1 and c-Ets-2. Oncogene, 9***,** *911-21.*

*COHEN, M. H., GOOTENBERG, J., KEEGAN, P. & PAZDUR, R. 2007. FDA Drug Approval Summary: Bevacizumab (Avastin(R)) Plus Carboplatin and Paclitaxel as First-Line Treatment of Advanced/Metastatic Recurrent Nonsquamous Non-Small Cell Lung Cancer. Oncologist, 12***,** *713-718.*

*CONFORTI, F., SAYAN, A. E., SREEKUMAR, R. & SAYAN, B. S. 2012. Regulation of p73 activity by post-translational modifications. Cell Death Dis, 3***,** *e285.*

*CONWAY, T. & SCHOOLNIK, K. 2003. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. Molecular Microbiology, 47***,** *879-889.*

*COOK, A. C., CHAMBERS, A. F., TURLEY, E. A. & TUCK, A. B. 2006. Osteopontin induction of hyaluronan synthase 2 expression promotes breast cancer malignancy. J Biol Chem, 281***,** *24381-9.*

*CORONA, R. 1996. Epidemiology of nonmelanoma skin cancer: a review. Ann Ist Super Sanita, 32***,** *37-42.*

*COSKUN, V., TSOA, R. & SUN, Y. E. 2012. Epigenetic regulation of stem cells differentiating along the neural lineage. Curr Opin Neurobiol, 22***,** *762-7.*

*COUSSENS, L. M. & WERB, Z. 2002. Inflammation and cancer. Nature, 420***,** *860-7.*

*COWEY, C. L. 2013. Targeted Therapy for Advanced Basal-Cell Carcinoma: Vismodegib and Beyond. Dermatology and Therapy, 3***,** *17-31.*

*CRAIG, D. H., DOWNEY, C. & BASSON, M. D. 2008. SiRNA-mediated reduction of alpha-actinin-1 inhibits pressure-induced murine tumor cell wound implantation and enhances tumor-free survival. Neoplasia, 10***,** *217-22.*

*CRETNIK, M., MUSANI, V., ORESKOVIC, S., LEOVIC, D. & LEVANAT, S. 2007. The Patched gene is epigenetically regulated in ovarian dermoids and fibromas, but not in basocellular carcinomas. International journal of molecular medicine, 19***,** *875-83.*

*CRISCIONE, V. D., WEINSTOCK, M. A., NAYLOR, M. F., LUQUE, C., EIDE, M. J., BINGHAM, S. F. & FOR THE DEPARTMENT OF VETERAN AFFAIRS TOPICAL TRETINOIN CHEMOPREVENTION TRIAL, G. 2009. Actinic keratoses. Cancer, 115***,** *2523-2530.*

*CROCE, C. M. 2009. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet, 10***,** *704-14.*

*CUI, Y., WANG, W., DONG, N., LOU, J., SRINIVASAN, D. K., CHENG, W., HUANG, X., LIU, M., FANG, C., PENG, J., CHEN, S., WU, S., LIU, Z., DONG, L., ZHOU, Y. & WU, Q. 2012. Role of corin in trophoblast invasion and uterine spiral artery remodelling in pregnancy. Nature, 484***,** *246-50.*

*DA SILVA MEIRELLES, L., CHAGASTELLES, P. C. & NARDI, N. B. 2006. Mesenchymal stem cells reside in virtually all post-natal organs and tissues. J Cell Sci, 119***,** *2204-13.*

*DAKUBO, G. D., JAKUPCIAK, J. P., BIRCH-MACHIN, M. A. & PARR, R. L. 2007. Clinical implications and utility of field cancerization. Cancer Cell Int, 7***,** *2.*

*DALLA-FAVERA, R., BREGNI, M., ERIKSON, J., PATTERSON, D., GALLO, R. C. & CROCE, C. M. 1982. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proceedings of the National Academy of Sciences of the United States of America, 79***,** *7824-7.*

DALLAGLIO, K., PETRACHI, T., MARCONI, A., TRUZZI, F., LOTTI, R., SALTARI, A., MORANDI, P., PUVIANI, M., MAIORANA, A., ROOP, D. R. & PINCELLI, C. 2013. Isolation and characterization of squamous cell carcinoma-derived stem-like cells: role in tumor formation. Int J Mol Sci, 14, 19540-55.

DANG, C. V. 2012. MYC on the path to cancer. Cell, 149, 22-35.

DAVID, A. R. & ZIMMERMAN, M. R. 2010. Cancer: an old disease, a new disease or something in between? Nat Rev Cancer, 10, 728-733.

DAWSON, M. A. & KOUZARIDES, T. 2012. Cancer epigenetics: from mechanism to therapy. Cell, 150, 12-27.

DAYA-GROSJEAN, L., ROBERT, C., DROUGARD, C., SUAREZ, H. & SARASIN, A. 1993. High Mutation Frequency in ras Genes of Skin Tumors Isolated from DNA Repair Deficient Xeroderma Pigmentosum Patients. Cancer Research, 53, 1625-1629.

DE HERTOG, S. A., WENSVEEN, C. A., BASTIAENS, M. T., KIELICH, C. J., BERKHOUT, M. J., WESTENDORP, R. G., VERMEER, B. J. & BOUWES BAVINCK, J. N. 2001. Relation between smoking and skin cancer. J Clin Oncol, 19, 231-8.

DE VRIES, E., VAN DE POLL-FRANSE, L. V., LOUWMAN, W. J., DE GRUIJL, F. R. & COEBERGH, J. W. 2005. Predictions of skin cancer incidence in the Netherlands up to 2015. Br J Dermatol, 152, 481-8.

DEEB, K. K., TRUMP, D. L. & JOHNSON, C. S. 2007. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. Nat Rev Cancer, 7, 684-700.

DEEB, S. S., BISSET, D. & FU, L. 2010. Epigenetic control of expression of the human L- and M-pigment genes. Ophthalmic Physiol Opt, 30, 446-53.

DEGENHARDT, Y. & LAMPKIN, T. 2010. Targeting Polo-like kinase in cancer therapy. Clin Cancer Res, 16, 384-9.

DEININGER, M., BUCHDUNGER, E. & DRUKER, B. J. 2005. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood, 105, 2640-2653.

DERYNCK, R., AKHURST, R. J. & BALMAIN, A. 2001. TGF-beta signaling in tumor suppression and cancer progression. Nat Genet, 29, 117-29.

DEVI, B., BEHERA, B., PATRO, S., PATTNAIK, S. S. & PUHAN, M. R. 2013. Gorlin syndrome. Indian J Dermatol, 58, 241.

DHAWAN, P. & RICHMOND, A. 2002. Role of CXCL1 in tumorigenesis of melanoma. J Leukoc Biol, 72, 9-18.

DHILLON, A. S., HAGAN, S., RATH, O. & KOLCH, W. 2007. MAP kinase signalling pathways in cancer. Oncogene, 26, 3279-90.

DI COSTANZO, A., FESTA, L., DUVERGER, O., VIVO, M., GUERRINI, L., LA MANTIA, G., MORASSO, M. I. & CALABRO, V. 2009. Homeodomain protein Dlx3 induces phosphorylation-dependent p63 degradation. Cell Cycle, 8, 1185-95.

DIAMANTOPOULOU, Z., KITSOU, P., MENASHI, S., COURTY, J. & KATSORIS, P. 2012. Loss of receptor protein tyrosine phosphatase beta/zeta (RPTPbeta/zeta) promotes prostate cancer metastasis. J Biol Chem, 287, 40339-49.

DIEHN, M., ALIZADEH, A. A. & BROWN, P. O. 2000. Examining the living genome in health and disease with DNA microarrays. JAMA, 283, 2298-9.

*DING, L., LEY, T. J., LARSON, D. E., MILLER, C. A., KOBOLDT, D. C., WELCH, J. S., RITCHEY, J. K., YOUNG, M. A., LAMPRECHT, T., MCLELLAN, M. D., MCMICHAEL, J. F., WALLIS, J. W., LU, C., SHEN, D., HARRIS, C. C., DOOLING, D. J., FULTON, R. S., FULTON, L. L., CHEN, K., SCHMIDT, H., KALICKI-VEIZER, J., MAGRINI, V. J., COOK, L., MCGRATH, S. D., VICKERY, T. L., WENDL, M. C., HEATH, S., WATSON, M. A., LINK, D. C., TOMASSON, M. H., SHANNON, W. D., PAYTON, J. E., KULKARNI, S., WESTERVELT, P., WALTER, M. J., GRAUBERT, T. A., MARDIS, E. R., WILSON, R. K. & DIPERSIO, J. F. 2012. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature, 481***,** *506-10.*

*DJOUAD, F., BONY, C., APPARAILLY, F., LOUIS-PLENCE, P., JORGENSEN, C. & NOEL, D. 2006. Earlier onset of syngeneic tumors in the presence of mesenchymal stem cells. Transplantation, 82***,** *1060 - 6.*

*DOBLER, M., SCHUH, J., KIESEWETTER, F., SCHELL, H., LIEHR, T. & GEBHART, E. 1999. Deletion monitoring in skin tumors by interphase-FISH using band-specific DNA probes. Int J Oncol, 14***,** *571-6.*

*DODD, L. E., SENGUPTA, S., CHEN, I. H., DEN BOON, J. A., CHENG, Y. J., WESTRA, W., NEWTON, M. A., MITTL, B. F., MCSHANE, L., CHEN, C. J., AHLQUIST, P. & HILDESHEIM, A. 2006. Genes involved in DNA repair and nitrosamine metabolism and those located on chromosome 14q32 are dysregulated in nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev, 15***,** *2216-25.*

DOLGIN, J. L. 2006. Surrounding Embryos: Biology, Ideology, and Politics. Health Matrix, 16.

*DREESEN, O. & BRIVANLOU, A. H. 2007. Signaling pathways in cancer and embryonic stem cells. Stem Cell Rev, 3***,** *7-17.*

*DRISKELL, I., ODA, H., BLANCO, S., NASCIMENTO, E., HUMPHREYS, P. & FRYE, M. 2012. The histone methyltransferase Setd8 acts in concert with c-Myc and is required to maintain skin. EMBO J, 31***,** *616-29.*

*DRUKKER, M., KATZ, G., URBACH, A., SCHULDINER, M., MARKEL, G., ITSKOVITZ-ELDOR, J., REUBINOFF, B., MANDELBOIM, O. & BENVENISTY, N. 2002. Characterization of the expression of MHC proteins in human embryonic stem cells. Proceedings of the National Academy of Sciences, 99***,** *9864-9869.*

*DU, P., KIBBE, W. A. & LIN, S. M. 2008. lumi: a pipeline for processing Illumina microarray. Bioinformatics, 24***,** *1547-8.*

*DUESBERG, P. H. & VOGT, P. K. 1970. Differences between the ribonucleic acids of transforming and nontransforming avian tumor viruses. Proceedings of the National Academy of Sciences of the United States of America, 67***,** *1673-80.*

*DUNCAVAGE, E. J., ZEHNBAUER, B. A. & PFEIFER, J. D. 2009. Prevalence of Merkel cell polyomavirus in Merkel cell carcinoma. Mod Pathol, 22***,** *516-21.*

*DUNNILL, M. G., RICHARDS, A. J., MILANA, G., MOLLICA, F., ATHERTON, D., WINSHIP, I., FARRALL, M., AL-IMARA, L., EADY, R. A. & POPE, F. M. 1994. Genetic linkage to the type VII collagen gene (COL7A1) in 26 families with generalised recessive dystrophic epidermolysis bullosa and anchoring fibril abnormalities. J Med Genet, 31***,** *745-8.*

*DVORAK, H. F. 1986. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N Engl J Med, 315***,** *1650-9.*

*DZIUNYCZ, P., IOTZOVA-WEISS, G., ELORANTA, J. J., LAUCHLI, S., HAFNER, J., FRENCH, L. E. & HOFBAUER, G. F. 2010. Squamous cell carcinoma of the skin shows a distinct microRNA profile modulated by UV radiation. J Invest Dermatol, 130***,** *2686-9.*

EICHBERGER, T., REGL, G., IKRAM, M. S., NEILL, G. W., PHILPOTT, M. P., ABERGER, F. & FRISCHAUF, A. M. 2004. FOXE1, a new transcriptional target of GLI2 is expressed in human epidermis and basal cell carcinoma. J Invest Dermatol, 122, 1180-7.

EISENBERG, I., ERAN, A., NISHINO, I., MOGGIO, M., LAMPERTI, C., AMATO, A. A., LIDOV, H. G., KANG, P. B., NORTH, K. N., MITRANI-ROSENBAUM, S., FLANIGAN, K. M., NEELY, L. A., WHITNEY, D., BEGGS, A. H., KOHANE, I. S. & KUNKEL, L. M. 2007. Distinctive patterns of microRNA expression in primary muscular disorders. Proc Natl Acad Sci U S A, 104, 17016-21.

ENERLY, E., STEINFELD, I., KLEIVI, K., LEIVONEN, S. K., AURE, M. R., RUSSNES, H. G., RONNEBERG, J. A., JOHNSEN, H., NAVON, R., RODLAND, E., MAKELA, R., NAUME, B., PERALA, M., KALLIONIEMI, O., KRISTENSEN, V. N., YAKHINI, Z. & BORRESEN-DALE, A. L. 2011. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. PLoS One, 6, e16915.

ENGLISH, C. M., ADKINS, M. W., CARSON, J. J., CHURCHILL, M. E. & TYLER, J. K. 2006. Structural basis for the histone chaperone activity of Asf1. Cell, 127, 495-508.

ESTELLER, M. 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet, 8, 286-98.

ESTELLER, M. 2008. Epigenetics in cancer. N Engl J Med, 358, 1148-59.

EVAN, G. I. & VOUSDEN, K. H. 2001. Proliferation, cell cycle and apoptosis in cancer. Nature, 411, 342-8.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature, 292, 154-156.

FAN, H. & JOHNSON, C. 2011. Insertional oncogenesis by non-acute retroviruses: implications for gene therapy. Viruses, 3, 398-422.

FANG, L., CHENG, J. C., CHANG, H. M., SUN, Y. P. & LEUNG, P. C. 2013. EGF-like Growth Factors Induce COX-2-Derived PGE2 Production Through ERK1/2 in Human Granulosa Cells. J Clin Endocrinol Metab, 98, 4932-41.

FASSAN, M., VOLINIA, S., PALATINI, J., PIZZI, M., BAFFA, R., DE BERNARD, M., BATTAGLIA, G., PARENTE, P., CROCE, C. M., ZANINOTTO, G., ANCONA, E. & RUGGE, M. 2011. MicroRNA expression profiling in human Barrett's carcinogenesis. Int J Cancer, 129, 1661-70.

FAVIER, B., ALAM, A., BARRON, P., BONNIN, J., LABOUDIE, P., FONS, P., MANDRON, M., HERAULT, J. P., NEUFELD, G., SAVI, P., HERBERT, J. M. & BONO, F. 2006. Neuropilin-2 interacts with VEGFR-2 and VEGFR-3 and promotes human endothelial cell survival and migration. Blood, 108, 1243-50.

FEBER, A., GUILHAMON, P., LECHNER, M., FENTON, T., WILSON, G. A., THIRLWELL, C., MORRIS, T. J., FLANAGAN, A. M., TESCHENDORFF, A. E., KELLY, J. D. & BECK, S. 2014. Using high-density DNA methylation arrays to profile copy number alterations. Genome Biol, 15, R30.

FENG, H., SHUDA, M., CHANG, Y. & MOORE, P. S. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science, 319, 1096-100.

FERLAY, J., AUTIER, P., BONIOL, M., HEANUE, M., COLOMBET, M. & BOYLE, P. 2007. Estimates of the cancer incidence and mortality in Europe in 2006. Annals of Oncology, 18, 581-592.

FICZ, G., BRANCO, M. R., SEISENBERGER, S., SANTOS, F., KRUEGER, F., HORE, T. A., MARQUES, C. J., ANDREWS, S. & REIK, W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature, 473, 398-402.

FIGUEROA, M. E., ABDEL-WAHAB, O., LU, C., WARD, P. S., PATEL, J., SHIH, A., LI, Y., BHAGWAT, N., VASANTHAKUMAR, A., FERNANDEZ, H. F., TALLMAN, M. S., SUN, Z., WOLNIAK, K., PEETERS, J. K., LIU, W., CHOE, S. E., FANTIN, V. R., PAIETTA, E., LOWENBERG, B., LICHT, J. D., GODLEY, L. A., DELWEL, R., VALK, P. J., THOMPSON, C. B., LEVINE, R. L. & MELNICK, A. 2010a. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell, 18, 553-67.

FIGUEROA, M. E., LUGTHART, S., LI, Y., ERPELINCK-VERSCHUEREN, C., DENG, X., CHRISTOS, P. J., SCHIFANO, E., BOOTH, J., VAN PUTTEN, W., SKRABANEK, L., CAMPAGNE, F., MAZUMDAR, M., GREALLY, J. M., VALK, P. J., LOWENBERG, B., DELWEL, R. & MELNICK, A. 2010b. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell, 17, 13-27.

FINE, J.-D., JOHNSON, L. B., WEINER, M., LI, K.-P. & SUCHINDRAN, C. 2009. Epidermolysis bullosa and the risk of life-threatening cancers: The National EB Registry experience, 1986-2006. Journal of the American Academy of Dermatology, 60, 203-211.

FINE, J. D., JOHNSON, L. B., WEINER, M. & SUCHINDRAN, C. 2008. Cause-specific risks of childhood death in inherited epidermolysis bullosa. J Pediatr, 152, 276-80.

FLEMING, J. L., DWORKIN, A. M., ALLAIN, D. C., FERNANDEZ, S., WEI, L., PETERS, S. B., IWENOFU, O. H., RIDD, K., BASTIAN, B. C. & TOLAND, A. E. 2013. Allele-specific imbalance mapping identifies HDAC9 as a candidate gene for cutaneous squamous cell carcinoma. Int J Cancer.

FONG, C. Y., CHAK, L. L., BISWAS, A., TAN, J. H., GAUTHAMAN, K., CHAN, W. K. & BONGSO, A. 2011. Human Wharton's jelly stem cells have unique transcriptome profiles compared to human embryonic stem cells and other mesenchymal stem cells. Stem Cell Rev, 7, 1-16.

FOULONGNE, V., KLUGER, N., DEREURE, O., BRIEU, N., GUILLOT, B. & SEGONDY, M. 2008. Merkel cell polyomavirus and Merkel cell carcinoma, France. Emerging infectious diseases, 14, 1491-3.

FREDERICK, J. P., LIBERATI, N. T., WADDELL, D. S., SHI, Y. & WANG, X. F. 2004. Transforming growth factor beta-mediated transcriptional repression of c-myc is dependent on direct binding of Smad3 to a novel repressive Smad binding element. Mol Cell Biol, 24, 2546-59.

FRIEDENSTEIN, A. J., CHAILAKHJAN, R. K. & LALYKINA, K. S. 1970. The development of fibroblast colonies in monolayer cultures of guinea-pig bone marrow and spleen cells. Cell Tissue Kinet, 3, 393-403.

FRIEDENSTEIN, A. J., CHAILAKHYAN, R. K., LATSINIK, N. V., PANASYUK, A. F. & KEILISS-BOROK, I. V. 1974. Stromal cells responsible for transferring the microenvironment of the hemopoietic tissues. Cloning in vitro and retransplantation in vivo. Transplantation, 17, 331-40.

FRIEND, S. H., BERNARDS, R., ROGELJ, S., WEINBERG, R. A., RAPAPORT, J. M., ALBERT, D. M. & DRYJA, T. P. 1986. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature, 323, 643-6.

FU, J., FONG, K., BELLACOSA, A., ROSS, E., APOSTOLOU, S., BASSI, D. E., JIN, F., ZHANG, J., CAIRNS, P., IBANEZ DE CACERES, I., BRAUNEWELL, K. H. & KLEIN-SZANTO, A. J. 2008. VILIP-1 downregulation in non-small cell lung carcinomas: mechanisms and prediction of survival. PLoS One, 3, e1698.

FUCHS, E. 2008. Skin stem cells: rising to the surface. The Journal of Cell Biology, 180, 273-284.

FUCHS, E. 2009. Finding one's niche in the skin. Cell Stem Cell, 4, 499-502.

GAGNON-BARTSCH, J. A. & SPEED, T. P. 2012. Using control genes to correct for unwanted variation in microarray data. Biostatistics, 13, 539-52.

GALLAGHER, R. P., BAJDIK, C. D., FINCHAM, S., HILL, G. B., KEEFE, A. R., COLDMAN, A. & MCLEAN, D. I. 1996. Chemical exposures, medical history, and risk of squamous and basal cell carcinoma of the skin. Cancer Epidemiology Biomarkers & Prevention, 5, 419-424.

GANDARILLAS, A. & WATT, F. M. 1997. c-Myc promotes differentiation of human epidermal stem cells. Genes Dev, 11, 2869-82.

GANEPOLA, G. A., RUTLEDGE, J. R., SUMAN, P., YIENGPRUKSAWAN, A. & CHANG, D. H. 2014. Novel blood-based microRNA biomarker panel for early diagnosis of pancreatic cancer. World J Gastrointest Oncol, 6, 22-33.

GANZENMUELLER, T., YAKUSHKO, Y., KLUBA, J., HENKE-GENDO, C., GUTZMER, R. & SCHULZ, T. F. 2012. Next-generation sequencing fails to identify human virus sequences in cutaneous squamous cell carcinoma. International Journal of Cancer, 131, E1173-E1179.

GARCIA, D. M., BAEK, D., SHIN, C., BELL, G. W., GRIMSON, A. & BARTEL, D. P. 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat Struct Mol Biol, 18, 1139-46.

GARNESKI, K. M., WARCOLA, A. H., FENG, Q., KIVIAT, N. B., LEONARD, J. H. & NGHIEM, P. 2009a. Merkel cell polyomavirus is more frequently present in North American than Australian Merkel cell carcinoma tumors. J Invest Dermatol, 129, 246-8.

GARNESKI, K. M., WARCOLA, A. H., FENG, Q., KIVIAT, N. B., LEONARD, J. H. & NGHIEM, P. 2009b. Merkel cell polyomavirus is more frequently present in North American than Australian Merkel cell carcinoma tumors. The Journal of investigative dermatology, 129, 246-8.

GARZA, L. A., YANG, C. C., ZHAO, T. L., BLATT, H. B., LEE, M., HE, H. L., STANTON, D. C., CARRASCO, L., SPIEGEL, J. H., TOBIAS, J. W. & COTSARELIS, G. 2011. Bald scalp in men with androgenetic alopecia retains hair follicle stem cells but lacks CD200-rich and CD34-positive hair follicle progenitor cells. Journal of Clinical Investigation, 121, 613-622.

GASTALDI, C., BERTERO, T., XU, N., BOURGET-PONZIO, I., LEBRIGAND, K., FOURRE, S., POPA, A., CARDOT-LECCIA, N., MENEGUZZI, G., SONKOLY, E., PIVARCSI, A., MARI, B., BARBRY, P., PONZIO, G. & REZZONICO, R. 2014. miR-193b/365a cluster controls progression of epidermal squamous cell carcinoma. Carcinogenesis, 35, 1110-20.

GEBHARDT, C., NEMETH, J., ANGEL, P. & HESS, J. 2006. S100A8 and S100A9 in inflammation and cancer. Biochem Pharmacol, 72, 1622-31.

GENTLEMAN, R. C., CAREY, V. J., BATES, D. M., BOLSTAD, B., DETTLING, M., DUDOIT, S., ELLIS, B., GAUTIER, L., GE, Y., GENTRY, J., HORNIK, K., HOTHORN, T., HUBER, W., IACUS, S., IRIZARRY, R., LEISCH, F., LI, C., MAECHLER, M., ROSSINI, A. J., SAWITZKI, G., SMITH, C., SMYTH, G., TIERNEY, L., YANG, J. Y. & ZHANG, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol, 5, R80.

GERLINGER, M., ROWAN, A. J., HORSWELL, S., LARKIN, J., ENDESFELDER, D., GRONROOS, E., MARTINEZ, P., MATTHEWS, N., STEWART, A., TARPEY, P., VARELA, I., PHILLIMORE, B., BEGUM, S., MCDONALD, N. Q., BUTLER, A., JONES, D., RAINE, K., LATIMER, C., SANTOS, C. R., NOHADANI, M., EKLUND, A. C., SPENCER-DENE, B., CLARK, G., PICKERING, L., STAMP, G., GORE, M., SZALLASI, Z., DOWNWARD, J., FUTREAL, P. A. & SWANTON, C. 2012. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. New England Journal of Medicine, 366, 883-892.

GHALI, L., WONG, S. T., GREEN, J., TIDMAN, N. & QUINN, A. G. 1999. Gli1 protein is expressed in basal cell carcinomas, outer root sheath keratinocytes and a subpopulation of *mesenchymal cells in normal human skin. The Journal of investigative dermatology, 113***,** *595-9.*

*GHOUSSAINI, M., FLETCHER, O., MICHAILIDOU, K., TURNBULL, C., SCHMIDT, M. K., DICKS, E., DENNIS, J., WANG, Q., HUMPHREYS, M. K., LUCCARINI, C., BAYNES, C., CONROY, D., MARANIAN, M., AHMED, S., DRIVER, K., JOHNSON, N., ORR, N., DOS SANTOS SILVA, I., WAISFISZ, Q., MEIJERS-HEIJBOER, H., UITTERLINDEN, A. G., RIVADENEIRA, F., NETHERLANDS COLLABORATIVE GROUP ON HEREDITARY, B., OVARIAN, C., HALL, P., CZENE, K., IRWANTO, A., LIU, J., NEVANLINNA, H., AITTOMAKI, K., BLOMQVIST, C., MEINDL, A., SCHMUTZLER, R. K., MULLER-MYHSOK, B., LICHTNER, P., CHANG-CLAUDE, J., HEIN, R., NICKELS, S., FLESCH-JANYS, D., TSIMIKLIS, H., MAKALIC, E., SCHMIDT, D., BUI, M., HOPPER, J. L., APICELLA, C., PARK, D. J., SOUTHEY, M., HUNTER, D. J., CHANOCK, S. J., BROEKS, A., VERHOEF, S., HOGERVORST, F. B., FASCHING, P. A., LUX, M. P., BECKMANN, M. W., EKICI, A. B., SAWYER, E., TOMLINSON, I., KERIN, M., MARME, F., SCHNEEWEISS, A., SOHN, C., BURWINKEL, B., GUENEL, P., TRUONG, T., CORDINA-DUVERGER, E., MENEGAUX, F., BOJESEN, S. E., NORDESTGAARD, B. G., NIELSEN, S. F., FLYGER, H., MILNE, R. L., ALONSO, M. R., GONZALEZ-NEIRA, A., BENITEZ, J., ANTON-CULVER, H., ZIOGAS, A., BERNSTEIN, L., DUR, C. C., BRENNER, H., MULLER, H., ARNDT, V., STEGMAIER, C., FAMILIAL BREAST CANCER, S., JUSTENHOVEN, C., BRAUCH, H., BRUNING, T., GENE ENVIRONMENT INTERACTION OF BREAST CANCER IN GERMANY, N., WANG-GOHRKE, S., EILBER, U., DORK, T., SCHURMANN, P., BREMER, M., HILLEMANNS, P., BOGDANOVA, N. V., ANTONENKOVA, N. N., ROGOV, Y. I., KARSTENS, J. H., BERMISHEVA, M., PROKOFIEVA, D., et al. 2012. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet, 44***,** *312-8.*

*GIRAUD, G., RAMQVIST, T., RAGNARSSON-OLDING, B. & DALIANIS, T. 2008. DNA from BK virus and JC virus and from KI, WU, and MC polyomaviruses as well as from simian virus 40 is not detected in non-UV-light-associated primary malignant melanomas of mucous membranes. Journal of clinical microbiology, 46***,** *3595-8.*

*GISSMANN, L. & HAUSEN, H. Z. 1980. Partial characterization of viral DNA from human genital warts (condylomata acuminata). International Journal of Cancer, 25***,** *605-609.*

*GLUD, M., ROSSING, M., HOTHER, C., HOLST, L., HASTRUP, N., NIELSEN, F. C., GNIADECKI, R. & DRZEWIECKI, K. T. 2010. Downregulation of miR-125b in metastatic cutaneous malignant melanoma. Melanoma research, 20***,** *479-84.*

*GOLDBERG, M., RUMMELT, C., LAERM, A., HELMBOLD, P., HOLBACH, L. M. & BALLHAUSEN, W. G. 2006. Epigenetic silencing contributes to frequent loss of the fragile histidine triad tumour suppressor in basal cell carcinomas. The British journal of dermatology, 155***,** *1154-8.*

*GOPPNER, D. & LEVERKUS, M. 2011. Basal cell carcinoma: from the molecular understanding of the pathogenesis to targeted therapy of progressive disease. J Skin Cancer, 2011***,** *650258.*

*GORDON, G. M. & DU, W. 2011. Conserved RB functions in development and tumor suppression. Protein Cell, 2***,** *864-78.*

*GORE, A., LI, Z., FUNG, H. L., YOUNG, J. E., AGARWAL, S., ANTOSIEWICZ-BOURGET, J., CANTO, I., GIORGETTI, A., ISRAEL, M. A., KISKINIS, E., LEE, J. H., LOH, Y. H., MANOS, P. D., MONTSERRAT, N., PANOPOULOS, A. D., RUIZ, S., WILBERT, M. L., YU, J., KIRKNESS, E. F., IZPISUA BELMONTE, J. C., ROSSI, D. J., THOMSON, J. A., EGGAN, K., DALEY, G. Q., GOLDSTEIN, L. S. & ZHANG, K. 2011. Somatic coding mutations in human induced pluripotent stem cells. Nature, 471***,** *63-7.*

GOTTESMAN, M. M. 2002. Mechanisms of cancer drug resistance. Annu Rev Med, 53, 615-27.

GRAZIANO, V. & DE LAURENZI, V. 2011. Role of p63 in cancer development. Biochim Biophys Acta, 1816, 57-66.

GREAVES, M. & MALEY, C. C. 2012. Clonal evolution in cancer. Nature, 481, 306-13.

GRESSNER, O., SCHILLING, T., LORENZ, K., SCHULZE SCHLEITHOFF, E., KOCH, A., SCHULZE-BERGKAMEN, H., LENA, A. M., CANDI, E., TERRINONI, A., CATANI, M. V., OREN, M., MELINO, G., KRAMMER, P. H., STREMMEL, W. & MULLER, M. 2005. TAp63alpha induces apoptosis by activating signaling via death receptors and mitochondria. EMBO J, 24, 2458-71.

GRODSTEIN, F., SPEIZER, F. E. & HUNTER, D. J. 1995. A prospective study of incident squamous cell carcinoma of the skin in the nurses' health study. J Natl Cancer Inst, 87, 1061-6.

GROGER, C. J., GRUBINGER, M., WALDHOR, T., VIERLINGER, K. & MIKULITS, W. 2012. Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. PLoS One, 7, e51136.

GRONNIGER, E., WEBER, B., HEIL, O., PETERS, N., STAB, F., WENCK, H., KORN, B., WINNEFELD, M. & LYKO, F. 2010. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. PLoS Genet, 6, e1000971.

GROOM, H. C. T. & BISHOP, K. N. 2012. The tale of xenotropic murine leukemia virus-related virus. Journal of General Virology, 93, 915-924.

GUAN, K., NAYERNIA, K., MAIER, L. S., WAGNER, S., DRESSEL, R., LEE, J. H., NOLTE, J., WOLF, F., LI, M., ENGEL, W. & HASENFUSS, G. 2006. Pluripotency of spermatogonial stem cells from adult mouse testis. Nature, 440, 1199-203.

GUSTAFSSON, B., HONKANIEMI, E., GOH, S., GIRAUD, G., FORESTIER, E., VON DOBELN, U., ALLANDER, T., DALIANIS, T. & BOGDANOVIC, G. 2012. KI, WU, and Merkel cell polyomavirus DNA was not detected in guthrie cards of children who later developed acute lymphoblastic leukemia. J Pediatr Hematol Oncol, 34, 364-7.

HAFNER, C., HOUBEN, R., BAEURLE, A., RITTER, C., SCHRAMA, D., LANDTHALER, M. & BECKER, J. C. 2012. Activation of the PI3K/AKT pathway in Merkel cell carcinoma. PloS one, 7, e31255.

HAMEETMAN, L., COMMANDEUR, S., BAVINCK, J. N., WISGERHOF, H. C., DE GRUIJL, F. R., WILLEMZE, R., MULLENDERS, L., TENSEN, C. P. & VRIELING, H. 2013. Molecular profiling of cutaneous squamous cell carcinomas and actinic keratoses from organ transplant recipients. BMC Cancer, 13, 58.

HAN, J., LEE, Y., YEOM, K. H., NAM, J. W., HEO, I., RHEE, J. K., SOHN, S. Y., CHO, Y., ZHANG, B. T. & KIM, V. N. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell, 125, 887-901.

HANAHAN, D. & WEINBERG, R. A. 2000. The Hallmarks of Cancer. Cell, 100, 57-70.

HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. Cell, 144, 646-74.

HANCOX, J. G., SHERIDAN, S. C., FELDMAN, S. R. & FLEISCHER, A. B., JR. 2004. Seasonal variation of dermatologic disease in the USA: a study of office visits from 1990 to 1998. Int J Dermatol, 43, 6-11.

HANDSCHEL, J., MULLER, D., DEPPRICH, R. A., OMMERBORN, M. A., KUBLER, N. R., NAUJOKS, C., REIFENBERGER, J., SCHAFER, K. L. & BRAUNSTEIN, S. 2010. The new polyomavirus

*(MCPyV) does not affect the clinical course in MCCs. International journal of oral and maxillofacial surgery, 39***,** *1086-90.*

*HANKS, B. A., HOLTZHAUSEN, A., EVANS, K. S., JAMIESON, R., GIMPEL, P., CAMPBELL, O. M., HECTOR-GREENE, M., SUN, L., TEWARI, A., GEORGE, A., STARR, M., NIXON, A., AUGUSTINE, C., BEASLEY, G., TYLER, D. S., OSADA, T., MORSE, M. A., LING, L., LYERLY, H. K. & BLOBE, G. C. 2013. Type III TGF-beta receptor downregulation generates an immunotolerant tumor microenvironment. J Clin Invest, 123***,** *3925-40.*

*HARMON, R. M., SIMPSON, C. L., JOHNSON, J. L., KOETSIER, J. L., DUBASH, A. D., NAJOR, N. A., SARIG, O., SPRECHER, E. & GREEN, K. J. 2013. Desmoglein-1/Erbin interaction suppresses ERK activation to support epidermal differentiation. J Clin Invest, 123***,** *1556-70.*

*HARRIS, J., SALASCHE, S. & HARRIS, R. 2001. Can internet-based continuing medical education improve physicians' skin cancer knowledge and skills? Journal of General Internal Medicine, 16***,** *50-56.*

*HARRIS, R. A., WANG, T., COARFA, C., NAGARAJAN, R. P., HONG, C., DOWNEY, S. L., JOHNSON, B. E., FOUSE, S. D., DELANEY, A., ZHAO, Y., OLSHEN, A., BALLINGER, T., ZHOU, X., FORSBERG, K. J., GU, J., ECHIPARE, L., O'GEEN, H., LISTER, R., PELIZZOLA, M., XI, Y., EPSTEIN, C. B., BERNSTEIN, B. E., HAWKINS, R. D., REN, B., CHUNG, W. Y., GU, H., BOCK, C., GNIRKE, A., ZHANG, M. Q., HAUSSLER, D., ECKER, J. R., LI, W., FARNHAM, P. J., WATERLAND, R. A., MEISSNER, A., MARRA, M. A., HIRST, M., MILOSAVLJEVIC, A. & COSTELLO, J. F. 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol, 28***,** *1097-105.*

*HARVEY, I., FRANKEL, S., MARKS, R., SHALOM, D. & NOLAN-FARRELL, M. 1996. Non-melanoma skin cancer and solar keratoses. I. Methods and descriptive results of the South Wales Skin Cancer Study. Br J Cancer, 74***,** *1302-7.*

*HARWOOD, C. A. & PROBY, C. M. 2002. Human papillomaviruses and non-melanoma skin cancer. Current Opinion in Infectious Diseases, 15***,** *101-114.*

*HARWOOD, C. A., SURENTHERAN, T., MCGREGOR, J. M., SPINK, P. J., LEIGH, I. M., BREUER, J. & PROBY, C. M. 2000. Human papillomavirus infection and non-melanoma skin cancer in immunosuppressed and immunocompetent individuals. J Med Virol, 61***,** *289-97.*

*HASHIDA, Y., IMAJOH, M., NEMOTO, Y., KAMIOKA, M., TANIGUCHI, A., TAGUCHI, T., KUME, M., ORIHASHI, K. & DAIBATA, M. 2013. Detection of Merkel cell polyomavirus with a tumour-specific signature in non-small cell lung cancer. Br J Cancer, 108***,** *629-637.*

*HE, B., CAUDY, A., PARSONS, L., ROSEBROCK, A., PANE, A., RAJ, S. & WIESCHAUS, E. 2012. Mapping the pericentric heterochromatin by comparative genomic hybridization analysis and chromosome deletions in Drosophila melanogaster. Genome Res, 22***,** *2507-19.*

*HEDMAN, H., NILSSON, J., GUO, D. & HENRIKSSON, R. 2002. Is LRIG1 a tumour suppressor gene at chromosome 3p14.3? Acta Oncol, 41***,** *352-4.*

*HEFFELFINGER, C., OUYANG, Z., ENGBERG, A., LEFFELL, D. J., HANLON, A. M., GORDON, P. B., ZHENG, W., ZHAO, H., SNYDER, M. P. & BALE, A. E. 2012. Correlation of Global MicroRNA Expression With Basal Cell Carcinoma Subtype. G3 (Bethesda), 2***,** *279-86.*

*HEIJMANS, B. T., TOBI, E. W., STEIN, A. D., PUTTER, H., BLAUW, G. J., SUSSER, E. S., SLAGBOOM, P. E. & LUMEY, L. H. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proceedings of the National Academy of Sciences, 105***,** *17046-17049.*

HEINRICHS, S., LI, C. & LOOK, A. T. 2010. SNP array analysis in hematologic malignancies: avoiding false discoveries. Blood, 115, 4157-4161.

HEITZER, E., BAMBACH, I., DANDACHI, N., HORN, M. & WOLF, P. 2010. PTCH promoter methylation at low level in sporadic basal cell carcinoma analysed by three different approaches. Experimental dermatology, 19, 926-8.

HELIN, K. & DHANAK, D. 2013. Chromatin proteins and modifications as drug targets. Nature, 502, 480-488.

HELMBOLD, P., LAHTZ, C., ENK, A., HERRMANN-TROST, P., MARSCH, W., KUTZNER, H. & DAMMANN, R. H. 2009a. Frequent occurrence of RASSF1A promoter hypermethylation and Merkel cell polyomavirus in Merkel cell carcinoma. Molecular carcinogenesis, 48, 903-9.

HELMBOLD, P., LAHTZ, C., HERPEL, E., SCHNABEL, P. A. & DAMMANN, R. H. 2009b. Frequent hypermethylation of RASSF1A tumour suppressor gene promoter and presence of Merkel cell polyomavirus in small cell lung cancer. European journal of cancer, 45, 2207-11.

HENKEN, F. E., WILTING, S. M., OVERMEER, R. M., VAN RIETSCHOTEN, J. G., NYGREN, A. O., ERRAMI, A., SCHOUTEN, J. P., MEIJER, C. J., SNIJDERS, P. J. & STEENBERGEN, R. D. 2007. Sequential gene promoter methylation during HPV-induced cervical carcinogenesis. British journal of cancer, 97, 1457-64.

HENRICHSEN, C. N., CHAIGNAT, E. & REYMOND, A. 2009. Copy number variants, diseases and gene expression. Human Molecular Genetics, 18, R1-R8.

HERMAN, J. G. & BAYLIN, S. B. 2003. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med, 349, 2042-54.

HERMAN, J. G., MERLO, A., MAO, L., LAPIDUS, R. G., ISSA, J.-P. J., DAVIDSON, N. E., SIDRANSKY, D. & BAYLIN, S. B. 1995. Inactivation of the CDKN2/p16/MTS1 Gene Is Frequently Associated with Aberrant DNA Methylation in All Common Human Cancers. Cancer Research, 55, 4525-4530.

HILL, M. J., HAWKSWORTH, G. & TATTERSALL, G. 1973. Bacteria, nitrosamines and cancer of the stomach. Br J Cancer, 28, 562-7.

HIRANO, M., HASHIMOTO, S., YONEMURA, S., SABE, H. & AIZAWA, S. 2008. EPB41L5 functions to post-transcriptionally regulate cadherin and integrin during epithelial-mesenchymal transition. J Cell Biol, 182, 1217-30.

HIYAMA, K., KODAIRA, M. & SATOH, C. 1990. Detection of deletions, insertions and single nucleotide substitutions in cloned beta-globin genes and new polymorphic nucleotide substitutions in beta-globin genes in a Japanese population using ribonuclease cleavage at mismatches in RNA:DNA duplexes. Mutat Res, 231, 219-31.

HLAVANDA, E., KLEMENT, E., KOKAI, E., KOVACS, J., VINCZE, O., TOKESI, N., OROSZ, F., MEDZIHRADSZKY, K. F., DOMBRADI, V. & OVADI, J. 2007. Phosphorylation blocks the activity of tubulin polymerization-promoting protein (TPPP): identification of sites targeted by different kinases. J Biol Chem, 282, 29531-9.

HOCKING, A. M. & GIBRAN, N. S. 2010. Mesenchymal stem cells: paracrine signaling and differentiation during cutaneous wound repair. Experimental cell research, 316, 2213-9.

HODGSON, N. C. 2005. Merkel cell carcinoma: changing incidence trends. Journal of surgical oncology, 89, 1-4.

*HOI, C. S., LEE, S. E., LU, S. Y., MCDERMITT, D. J., OSORIO, K. M., PISKUN, C. M., PETERS, R. M., PAUS, R. & TUMBAR, T. 2010. Runx1 directly promotes proliferation of hair follicle stem cells and epithelial tumor formation in mouse skin. Mol Cell Biol, 30***,** *2518-36.*

*HOLLWAY, G. E., MAULE, J., GAUTIER, P., EVANS, T. M., KEENAN, D. G., LOHS, C., FISCHER, D., WICKING, C. & CURRIE, P. D. 2006. Scube2 mediates Hedgehog signalling in the zebrafish embryo. Dev Biol, 294***,** *104-18.*

*HON, G. C., HAWKINS, R. D., CABALLERO, O. L., LO, C., LISTER, R., PELIZZOLA, M., VALSESIA, A., YE, Z., KUAN, S., EDSALL, L. E., CAMARGO, A. A., STEVENSON, B. J., ECKER, J. R., BAFNA, V., STRAUSBERG, R. L., SIMPSON, A. J. & REN, B. 2012. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res, 22***,** *246-58.*

*HONDA, N., JINNIN, M., KAJIHARA, I., MAKINO, T., MAKINO, K., MASUGUCHI, S., FUKUSHIMA, S., OKAMOTO, Y., HASEGAWA, M., FUJIMOTO, M. & IHN, H. 2012. TGF-beta-mediated downregulation of microRNA-196a contributes to the constitutive upregulated type I collagen expression in scleroderma dermal fibroblasts. J Immunol, 188***,** *3323-31.*

*HONG, E. E., OKITSU, C. Y., SMITH, A. D. & HSIEH, C.-L. 2013. Regionally Specific and Genome-Wide Analyses Conclusively Demonstrate the Absence of CpG Methylation in Human Mitochondrial DNA. Molecular and Cellular Biology, 33***,** *2683-2690.*

*HOUBEN, R., SHUDA, M., WEINKAM, R., SCHRAMA, D., FENG, H., CHANG, Y., MOORE, P. S. & BECKER, J. C. 2010. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. Journal of virology, 84***,** *7064-72.*

*HOUSEMAN, E. A., CHRISTENSEN, B. C., KARAGAS, M. R., WRENSCH, M. R., NELSON, H. H., WIEMELS, J. L., ZHENG, S., WIENCKE, J. K., KELSEY, K. T. & MARSIT, C. J. 2009. Copy number variation has little impact on bead-array-based measures of DNA methylation. Bioinformatics, 25***,** *1999-2005.*

*HOUSMAN, T. S., FELDMAN, S. R., WILLIFORD, P. M., FLEISCHER, A. B., JR., GOLDMAN, N. D., ACOSTAMADIEDO, J. M. & CHEN, G. J. 2003. Skin cancer is among the most costly of all cancers to treat for the Medicare population. J Am Acad Dermatol, 48***,** *425-9.*

*HOVE, J. R., KOSTER, R. W., FOROUHAR, A. S., ACEVEDO-BOLTON, G., FRASER, S. E. & GHARIB, M. 2003. Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis. Nature, 421***,** *172-7.*

*HOVNANIAN, A., DUQUESNOY, P., BLANCHET-BARDON, C., KNOWLTON, R. G., AMSELEM, S., LATHROP, M., DUBERTRET, L., UITTO, J. & GOOSSENS, M. 1992. Genetic linkage of recessive dystrophic epidermolysis bullosa to the type VII collagen gene. J Clin Invest, 90***,** *1032-6.*

*HU, C. Y., MOHTAT, D., YU, Y., KO, Y. A., SHENOY, N., BHATTACHARYA, S., IZQUIERDO, M. C., PARK, A. S., GIRICZ, O., VALLUMSETLA, N., GUNDABOLU, K., WARE, K., BHAGAT, T. D., SUZUKI, M., PULLMAN, J., LIU, X. S., GREALLY, J. M., SUSZTAK, K. & VERMA, A. 2014. Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. Clin Cancer Res, 20***,** *4349-60.*

*HUANG, Y., GUIGON, C. J., FAN, J., CHENG, S. Y. & ZHU, G. Z. 2010. Pituitary homeobox 2 (PITX2) promotes thyroid carcinogenesis by activation of cyclin D2. Cell Cycle, 9***,** *1333-41.*

*HUMPHRIES, J. D., BYRON, A., BASS, M. D., CRAIG, S. E., PINNEY, J. W., KNIGHT, D. & HUMPHRIES, M. J. 2009. Proteomic analysis of integrin-associated complexes identifies RCC2 as a dual regulator of Rac1 and Arf6. Sci Signal, 2***,** *ra51.*

*HUNG, P. S., KAO, S. Y., SHIH, Y. H., CHIOU, S. H., LIU, C. J., CHANG, K. W. & LIN, S. C. 2008. Insulin-like growth factor binding protein-5 (IGFBP-5) suppresses the tumourigenesis of head and neck squamous cell carcinoma. J Pathol, 214***,** *368-76.*

*HUSSEIN, S. M., BATADA, N. N., VUORISTO, S., CHING, R. W., AUTIO, R., NARVA, E., NG, S., SOUROUR, M., HAMALAINEN, R., OLSSON, C., LUNDIN, K., MIKKOLA, M., TROKOVIC, R., PEITZ, M., BRUSTLE, O., BAZETT-JONES, D. P., ALITALO, K., LAHESMAA, R., NAGY, A. & OTONKOSKI, T. 2011. Copy number variation and selection during reprogramming to pluripotency. Nature, 471***,** *58-62.*

*IBANEZ-VENTOSO, C., VORA, M. & DRISCOLL, M. 2008. Sequence relationships among C. elegans, D. melanogaster and human microRNAs highlight the extensive conservation of microRNAs in biology. PLoS One, 3***,** *e2818.*

*IBIEBELE, T. I., VAN DER POLS, J. C., HUGHES, M. C., MARKS, G. C., WILLIAMS, G. M. & GREEN, A. C. 2007. Dietary pattern in association with squamous cell carcinoma of the skin: a prospective study. The American Journal of Clinical Nutrition, 85***,** *1401-1408.*

*IBRAGIMOVA, I., SLIFKER, M. J., MARADEO, M. E., BANUMATHY, G., DULAIMI, E., UZZO, R. G. & CAIRNS, P. 2013. Genome-wide promoter methylome of small renal masses. PLoS One, 8***,** *e77309.*

*ICHIHARA, A., JINNIN, M., YAMANE, K., FUJISAWA, A., SAKAI, K., MASUGUCHI, S., FUKUSHIMA, S., MARUO, K. & IHN, H. 2011. microRNA-mediated keratinocyte hyperproliferation in psoriasis vulgaris. Br J Dermatol, 165***,** *1003-10.*

*ICHIHARA, A., WANG, Z., JINNIN, M., IZUNO, Y., SHIMOZONO, N., YAMANE, K., FUJISAWA, A., MORIYA, C., FUKUSHIMA, S., INOUE, Y. & IHN, H. 2014. Upregulation of miR-18a-5p contributes to epidermal necrolysis in severe drug eruptions. J Allergy Clin Immunol, 133***,** *1065-74.*

*IGLESIAS-BARTOLOME, R., CALLEJAS-VALERA, J. L. & GUTKIND, J. S. 2013. Control of the epithelial stem cell epigenome: the shaping of epithelial stem cell identity. Current Opinion in Cell Biology, 25***,** *162-169.*

*IKEHATA, H., HIGASHI, S., NAKAMURA, S., DAIGAKU, Y., FURUSAWA, Y., KAMEI, Y., WATANABE, M., YAMAMOTO, K., HIEDA, K., MUNAKATA, N. & ONO, T. 2013. Action spectrum analysis of UVR genotoxicity for skin: the border wavelengths between UVA and UVB can bring serious mutation loads to skin. J Invest Dermatol, 133***,** *1850-6.*

*INOUE, A. & ZHANG, Y. 2011. Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos. Science, 334***,** *194.*

*INOUE, K., JINNIN, M., YAMANE, K., MAKINO, T., KAJIHARA, I., MAKINO, K., HONDA, N., NAKAYAMA, W., FUKUSHIMA, S. & IHN, H. 2013. Down-regulation of miR-223 contributes to the formation of Gottron's papules in dermatomyositis via the induction of PKCvarepsilon. Eur J Dermatol, 23***,** *160-7.*

*IRIZARRY, R. A., LADD-ACOSTA, C., WEN, B., WU, Z., MONTANO, C., ONYANGO, P., CUI, H., GABO, K., RONGIONE, M., WEBSTER, M., JI, H., POTASH, J. B., SABUNCIYAN, S. & FEINBERG, A. P. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet, 41***,** *178-86.*

*ITO, S., D'ALESSIO, A. C., TARANOVA, O. V., HONG, K., SOWERS, L. C. & ZHANG, Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature, 466***,** *1129-33.*

ITO, Y., KUROKAWA, I., NISHIMURA, K., HAKAMADA, A., ISODA, K., YAMANAKA, K., TSUBURA, A. & MIZUTANI, H. 2008. Keratin and filaggrin expression in keratoacanthoma. J Eur Acad Dermatol Venereol, 22, 353-5.

JACKSON, S. J., ZHANG, Z., FENG, D., FLAGG, M., O'LOUGHLIN, E., WANG, D., STOKES, N., FUCHS, E. & YI, R. 2013. Rapid and widespread suppression of self-renewal by microRNA-203 during epidermal differentiation. Development, 140, 1882-91.

JEMAL, A., SIEGEL, R., WARD, E., MURRAY, T., XU, J., SMIGAL, C. & THUN, M. J. 2006. Cancer Statistics, 2006. CA Cancer J Clin, 56, 106-130.

JENSEN, K. B., COLLINS, C. A., NASCIMENTO, E., TAN, D. W., FRYE, M., ITAMI, S. & WATT, F. M. 2009. Lrig1 expression defines a distinct multipotent stem cell population in mammalian epidermis. Cell Stem Cell, 4, 427-39.

JEONG, J., MAO, J., TENZEN, T., KOTTMANN, A. H. & MCMAHON, A. P. 2004. Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. Genes Dev, 18, 937-51.

JIN, Y., MARTINS, C., JIN, C., SALEMARK, L., JONSSON, N., PERSSON, B., ROQUE, L., FONSECA, I. & WENNERBERG, J. 1999. Nonrandom karyotypic features in squamous cell carcinomas of the skin. Genes Chromosomes Cancer, 26, 295-303.

JINNIN, M. 2014. Various applications of microRNAs in skin diseases. J Dermatol Sci, 74, 3-8.

JOHNE, R., BUCK, C., ALLANDER, T., ATWOOD, W., GARCEA, R., IMPERIALE, M., MAJOR, E., RAMQVIST, T. & NORKIN, L. 2011. Taxonomical developments in the family Polyomaviridae. Archives of Virology, 156, 1627-1634.

JOHNSON, T. M., ROWE, D. E., NELSON, B. R. & SWANSON, N. A. 1992. Squamous cell carcinoma of the skin (excluding lip and oral mucosa). Journal of the American Academy of Dermatology, 26, 467-484.

JONES, P. A. & TAKAI, D. 2001. The role of DNA methylation in mammalian epigenetics. Science, 293, 1068-70.

JONES, R. G. & THOMPSON, C. B. 2009. Tumor suppressors and cell metabolism: a recipe for cancer growth. Genes & Development, 23, 537-548.

JOYCE, C. E., ZHOU, X., XIA, J., RYAN, C., THRASH, B., MENTER, A., ZHANG, W. & BOWCOCK, A. M. 2011. Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. Human Molecular Genetics, 20, 4025-4040.

JUNG, E. M., SON, H. Y., JEUNG, M. K., LEE, C. K., HYUN, S. H. & JEUNG, E. B. 2011. Epigenetic signatures of somatic cell nuclear transfer-derived embryonic stem cells. Int J Mol Med, 28, 697-704.

JUNG, M., SCHAEFER, A., STEINER, I., KEMPKENSTEFFEN, C., STEPHAN, C., ERBERSDOBLER, A. & JUNG, K. 2010. Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. Clin Chem, 56, 998-1006.

JURAK, I., KRAMER, M. F., MELLOR, J. C., VAN LINT, A. L., ROTH, F. P., KNIPE, D. M. & COEN, D. M. 2010. Numerous conserved and divergent microRNAs expressed by herpes simplex viruses 1 and 2. J Virol, 84, 4659-72.

JURIC, D., LACAYO, N. J., RAMSEY, M. C., RACEVSKIS, J., WIERNIK, P. H., ROWE, J. M., GOLDSTONE, A. H., O'DWYER, P. J., PAIETTA, E. & SIKIC, B. I. 2007. Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias. J Clin Oncol, 25, 1341-9.

*KADANGA, A. K., LEROUX, C., BONNET, M., CHAUVET, S., MEUNIER, B., CASSAR-MALEK, I. & HOCQUETTE, J. F. 2008. Image analysis and data normalization procedures are crucial for microarray analyses. Gene Regul Syst Bio, 2***,** *107-12.*

*KAMINSKAS, E., FARRELL, A., ABRAHAM, S., BAIRD, A., HSIEH, L.-S., LEE, S.-L., LEIGHTON, J. K., PATEL, H., RAHMAN, A., SRIDHARA, R., WANG, Y.-C. & PAZDUR, R. 2005. Approval Summary: Azacitidine for Treatment of Myelodysplastic Syndrome Subtypes. Clinical Cancer Research, 11***,** *3604-3608.*

*KANG, Y. & MASSAGUE, J. 2004. Epithelial-mesenchymal transitions: twist in development and metastasis. Cell, 118***,** *277-9.*

*KANITZ, A., IMIG, J., DZIUNYCZ, P. J., PRIMORAC, A., GALGANO, A., HOFBAUER, G. F., GERBER, A. P. & DETMAR, M. 2012. The expression levels of microRNA-361-5p and its target VEGFA are inversely correlated in human cutaneous squamous cell carcinoma. PLoS One, 7***,** *e49568.*

*KANJILAL, S., STROM, S. S., CLAYMAN, G. L., WEBER, R. S., EL-NAGGAR, A. K., KAPUR, V., CUMMINGS, K. K., HILL, L. A., SPITZ, M. R., KRIPKE, M. L. & ET AL. 1995. p53 mutations in nonmelanoma skin cancer of the head and neck: molecular evidence for field cancerization. Cancer Res, 55***,** *3604-9.*

*KARAGAS, M. R., NELSON, H. H., SEHR, P., WATERBOER, T., STUKEL, T. A., ANDREW, A., GREEN, A. C., BAVINCK, J. N., PERRY, A., SPENCER, S., REES, J. R., MOTT, L. A. & PAWLITA, M. 2006. Human papillomavirus infection and incidence of squamous cell and basal cell carcinomas of the skin. J Natl Cancer Inst, 98***,** *389-95.*

*KARAGAS, M. R., STUKEL, T. A., GREENBERG, E. R., BARON, J. A., MOTT, L. A. & STERN, R. S. 1992. Risk of subsequent basal cell carcinoma and squamous cell carcinoma of the skin among patients with prior skin cancer. Skin Cancer Prevention Study Group. JAMA, 267***,** *3305-10.*

*KARAMBOULAS, C. & AILLES, L. 2013. Developmental signaling pathways in cancer stem cells of solid tumors. Biochimica et Biophysica Acta (BBA) - General Subjects, 1830***,** *2481-2495.*

*KARNOUB, A., DASH, A., VO, A., SULLIVAN, A., BROOKS, M., BELL, G., RICHARDSON, A., POLYAK, K., TUBO, R. & WEINBERG, R. 2007. Mesenchymal stem cells within tumor stroma promote breast cancer metastasis. Nature, 449***,** *557 - 63.*

*KASSEM, A., TECHNAU, K., KURZ, A. K., PANTULU, D., LONING, M., KAYSER, G., STICKELER, E., WEYERS, W., DIAZ, C., WERNER, M., NASHAN, D. & ZUR HAUSEN, A. 2009. Merkel cell polyomavirus sequences are frequently detected in nonmelanoma skin cancer of immunosuppressed patients. Int J Cancer, 125***,** *356-61.*

*KEAN, J. M., RAO, S., WANG, M. & GARCEA, R. L. 2009. Seroepidemiology of human polyomaviruses. PLoS pathogens, 5***,** *e1000363.*

*KENNEDY CRISPIN, M., FUENTES-DUCULAN, J., GULATI, N., JOHNSON-HUANG, L. M., LENTINI, T., SULLIVAN-WHALEN, M., GILLEAUDEAU, P., CUETO, I., SUAREZ-FARINAS, M., LOWES, M. A. & KRUEGER, J. G. 2013. Gene Profiling of Narrowband UVB-Induced Skin Injury Defines Cellular and Molecular Innate Immune Responses. J Invest Dermatol, 133***,** *692-701.*

*KHAKOO, A., PATI, S., ANDERSON, S., REID, W., ELSHAL, M., ROVIRA, I., NGUYEN, A., MALIDE, D., COMBS, C., HALL, G., ZHANG, J., RAFFELD, M., ROGERS, T., STETLER-STEVENSON, W., FRANK, J., REITZ, M. & FINKEL, T. 2006. Human mesenchymal stem cells exert potent antitumorigenic effects in a model of Kaposi's sarcoma. J Exp Med, 203***,** *1235 - 47.*

*KIM, M., KANG, T. W., LEE, H. C., HAN, Y. M., KIM, H., SHIN, H. D., CHEONG, H. S., LEE, D., KIM, S. Y. & KIM, Y. S. 2011. Identification of DNA methylation markers for lineage commitment of in vitro hepatogenesis. Hum Mol Genet, 20***,** *2722-33.*

*KITKUMTHORN, N., YANATATSANAJIT, P., KIATPONGSAN, S., PHOKAEW, C., TRIRATANACHAT, S., TRIVIJITSILP, P., TERMRUNGRUANGLERT, W., TRESUKOSOL, D., NIRUTHISARD, S. & MUTIRANGURA, A. 2006. Cyclin A1 promoter hypermethylation in human papillomavirus-associated cervical cancer. BMC cancer, 6***,** *55.*

*KLEINSMITH, L. J. & PIERCE, G. B. 1964. Multipotentiality of Single Embryonal Carcinoma Cells. Cancer Research, 24***,** *1544-1551.*

*KNOEPFLER, P. S. 2009. Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine. Stem Cells, 27***,** *1050-6.*

*KOINUMA, D., TSUTSUMI, S., KAMIMURA, N., TANIGUCHI, H., MIYAZAWA, K., SUNAMURA, M., IMAMURA, T., MIYAZONO, K. & ABURATANI, H. 2009. Chromatin immunoprecipitation on microarray analysis of Smad2/3 binding sites reveals roles of ETS1 and TFAP2A in transforming growth factor beta signaling. Mol Cell Biol, 29***,** *172-86.*

*KOLASINSKA-ZWIERZ, P., DOWN, T., LATORRE, I., LIU, T., LIU, X. S. & AHRINGER, J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet, 41***,** *376-81.*

*KOLJONEN, V., KUKKO, H., PUKKALA, E., SANKILA, R., BOHLING, T., TUKIAINEN, E., SIHTO, H. & JOENSUU, H. 2009. Chronic lymphocytic leukaemia patients have a high risk of Merkel-cell polyomavirus DNA-positive Merkel-cell carcinoma. British journal of cancer, 101***,** *1444-7.*

*KONG, B., MICHALSKI, C. W., HONG, X., VALKOVSKAYA, N., RIEDER, S., ABIATARI, I., STREIT, S., ERKAN, M., ESPOSITO, I., FRIESS, H. & KLEEFF, J. 2010. AZGP1 is a tumor suppressor in pancreatic cancer inducing mesenchymal-to-epithelial transdifferentiation by inhibiting TGF-beta-mediated ERK signaling. Oncogene, 29***,** *5146-58.*

*KOSLOWSKI, M., LUXEMBURGER, U., TURECI, O. & SAHIN, U. 2011. Tumor-associated CpG demethylation augments hypoxia-induced effects by positive autoregulation of HIF-1alpha. Oncogene, 30***,** *876-82.*

KOSS, L. G. & MELAMED, M. R. 2006. Koss' diagnostic cytology and its histopathologic bases, Philadelphia, Lippincott Williams & Wilkins.

*KOSTER, M. E. & BERGSMA, J. 1990. Problems and coping behaviour of facial cancer patients. Soc Sci Med, 30***,** *569-78.*

*KOSTER, M. I., KIM, S., MILLS, A. A., DEMAYO, F. J. & ROOP, D. R. 2004. p63 is the molecular switch for initiation of an epithelial stratification program. Genes Dev, 18***,** *126-31.*

*KOU, G., SHI, J., CHEN, L., ZHANG, D., HOU, S., ZHAO, L., FANG, C., ZHENG, L., ZHANG, X., LIANG, P., LI, B. & GUO, Y. 2010. A bispecific antibody effectively inhibits tumor growth and metastasis by simultaneous blocking vascular endothelial growth factor A and osteopontin. Cancer Lett, 299***,** *130-6.*

*KOUIDOU, S., AGIDOU, T., KYRKOU, A., ANDREOU, A., KATOPODI, T., GEORGIOU, E., KRIKELIS, D., DIMITRIADOU, A., SPANOS, P., TSILIKAS, C., DESTOUNI, H. & TZIMAGIORGIS, G. 2005. Non-CpG cytosine methylation of p53 exon 5 in non-small cell lung carcinoma. Lung Cancer, 50***,** *299-307.*

KOZLENKOV, A., ROUSSOS, P., TIMASHPOLSKY, A., BARBU, M., RUDCHENKO, S., BIBIKOVA, M., KLOTZLE, B., BYNE, W., LYDDON, R., DI NARZO, A. F., HURD, Y. L., KOONIN, E. V. & DRACHEVA, S. 2013. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. Nucleic Acids Research.

KRECK, B., RICHTER, J., AMMERPOHL, O., BARANN, M., ESSER, D., PETERSEN, B. S., VATER, I., MURGA PENAS, E. M., BORMANN CHUNG, C. A., SEISENBERGER, S., LEE BOYD, V., SMALLWOOD, S., DREXLER, H. G., MACLEOD, R. A. F., HUMMEL, M., KRUEGER, F., HASLER, R., SCHREIBER, S., ROSENSTIEL, P., FRANKE, A. & SIEBERT, R. 2013. Base-pair resolution DNA methylome of the EBV-positive Endemic Burkitt lymphoma cell line DAUDI determined by SOLiD bisulfite-sequencing. Leukemia, 27, ***,*** *1751-1753.*

KRIAUCIONIS, S. & HEINTZ, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science, 324, *929-30.*

KRUEGER, F. & ANDREWS, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics, 27, *1571-2.*

KUBO, Y., URANO, Y., MATSUMOTO, K., AHSAN, K. & ARASE, S. 1997. Mutations of the INK4a locus in squamous cell carcinomas of human skin. Biochem Biophys Res Commun, 232, *38-41.*

KUKIMOTO, I., MORI, S., SATO, H., TAKEUCHI, T. & KANDA, T. 2008. Transcription factor human Skn-1a enhances replication of human papillomavirus DNA through the direct binding to two sites near the viral replication origin. FEBS J, 275, *3123-35.*

KULICH, M., RERICHA, V., RERICHA, R., SHORE, D. L. & SANDLER, D. P. 2011. Incidence of non-lung solid cancers in Czech uranium miners: a case-cohort study. Environ Res, 111, *400-5.*

KULIS, M., HEATH, S., BIBIKOVA, M., QUEIROS, A. C., NAVARRO, A., CLOT, G., MARTINEZ-TRILLOS, A., CASTELLANO, G., BRUN-HEATH, I., PINYOL, M., BARBERAN-SOLER, S., PAPASAIKAS, P., JARES, P., BEA, S., RICO, D., ECKER, S., RUBIO, M., ROYO, R., HO, V., KLOTZLE, B., HERNANDEZ, L., CONDE, L., LOPEZ-GUERRA, M., COLOMER, D., VILLAMOR, N., AYMERICH, M., ROZMAN, M., BAYES, M., GUT, M., GELPI, J. L., OROZCO, M., FAN, J. B., QUESADA, V., PUENTE, X. S., PISANO, D. G., VALENCIA, A., LOPEZ-GUILLERMO, A., GUT, I., LOPEZ-OTIN, C., CAMPO, E. & MARTIN-SUBERO, J. I. 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet, 44, *1236-42.*

KUNDAKOVIC, M., GUDSNUK, K., FRANKS, B., MADRID, J., MILLER, R. L., PERERA, F. P. & CHAMPAGNE, F. A. 2013. Sex-specific epigenetic disruption and behavioral changes following low-dose in utero bisphenol A exposure. Proceedings of the National Academy of Sciences, 110, *9956-9961.*

KYUNO, D., KOJIMA, T., YAMAGUCHI, H., ITO, T., KIMURA, Y., IMAMURA, M., TAKASAWA, A., MURATA, M., TANAKA, S., HIRATA, K. & SAWADA, N. 2013. Protein kinase Calpha inhibitor protects against downregulation of claudin-1 during epithelial-mesenchymal transition of pancreatic cancer. Carcinogenesis, 34, *1232-43.*

LAHTZ, C., KIM, S. I., BATES, S. E., LI, A. X., WU, X. & PFEIFER, G. P. 2013. UVB irradiation does not directly induce detectable changes of DNA methylation in human keratinocytes. F1000Res, 2, *45.*

LAI, D., VISSER-GRIEVE, S. & YANG, X. 2012. Tumour suppressor genes in chemotherapeutic drug response. Biosci Rep, 32, *361-74.*

LAIRD, P. W. 2003. The power and the promise of DNA methylation markers. Nat Rev Cancer, 3, *253-66.*

LAMBERT, S. R., HARWOOD, C. A., PURDIE, K. J., GULATI, A., MATIN, R. N., ROMANOWSKA, M., CERIO, R., KELSELL, D. P., LEIGH, I. M. & PROBY, C. M. 2012. Metastatic cutaneous squamous cell carcinoma shows frequent deletion in the protein tyrosine phosphatase receptor Type D gene. Int J Cancer, 131, *E216-26.*

*LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10***,** *R25.*

*LAPIDOT, T., SIRARD, C., VORMOOR, J., MURDOCH, B., HOANG, T., CACERES-CORTES, J., MINDEN, M., PATERSON, B., CALIGIURI, M. A. & DICK, J. E. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature, 367***,** *645-648.*

*LASITHIOTAKI, I., ANTONIOU, K. M., DERDAS, S. P., SARCHIANAKI, E., SYMVOULAKIS, E. K., PSARAKI, A., SPANDIDOS, D. A., STATHOPOULOS, E. N., SIAFAKAS, N. M. & SOURVINOS, G. 2013. The presence of Merkel cell polyomavirus is associated with deregulated expression of BRAF and Bcl-2 genes in non-small cell lung cancer. International Journal of Cancer, 133***,** *604-611.*

LAUDE, H. C., JONCHERE, B., MAUBEC, E., CARLOTTI, A., MARINHO, E., COUTURAUD, B., PETER, M., SASTRE-GARAU, X., AVRIL, M. F., DUPIN, N. & ROZENBERG, F. 2010. Distinct merkel cell polyomavirus molecular features in tumour and non tumour specimens from patients with merkel cell carcinoma. PLoS pathogens, 6.

*LAZEBNIK, Y. 2010. What are the hallmarks of cancer? Nat Rev Cancer, 10***,** *232-233.*

*LEAR, J. T., TAN, B. B., SMITH, A. G., JONES, P. W., HEAGERTY, A. H., STRANGE, R. C. & FRYER, A. A. 1998. A comparison of risk factors for malignant melanoma, squamous cell carcinoma and basal cell carcinoma in the UK. Int J Clin Pract, 52***,** *145-9.*

*LEE, H. S., LEE, D. C., PARK, M. H., YANG, S. J., LEE, J. J., KIM, D. M., JANG, Y., LEE, J. H., CHOI, J. Y., KANG, Y. K., KIM, D. I., PARK, K. C., KIM, S. Y., YOO, H. S., CHOI, E. J. & YEOM, Y. I. 2006a. STMN2 is a novel target of beta-catenin/TCF-mediated transcription in human hepatoma cells. Biochem Biophys Res Commun, 345***,** *1059-67.*

*LEE, J. H., AHN, K. S., LEE, C. H., YOUN, S. J., KIM, J. W., LEE, D. Y., LEE, E. S., STEINERT, P. M. & YANG, J. M. 2003a. Keratin 9 gene mutations in five Korean families with epidermolytic palmoplantar keratoderma. Exp Dermatol, 12***,** *876-81.*

*LEE, R. C., FEINBAUM, R. L. & AMBROS, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, 75***,** *843-54.*

*LEE, T. I., JENNER, R. G., BOYER, L. A., GUENTHER, M. G., LEVINE, S. S., KUMAR, R. M., CHEVALIER, B., JOHNSTONE, S. E., COLE, M. F., ISONO, K.-I., KOSEKI, H., FUCHIKAMI, T., ABE, K., MURRAY, H. L., ZUCKER, J. P., YUAN, B., BELL, G. W., HERBOLSHEIMER, E., HANNETT, N. M., SUN, K., ODOM, D. T., OTTE, A. P., VOLKERT, T. L., BARTEL, D. P., MELTON, D. A., GIFFORD, D. K., JAENISCH, R. & YOUNG, R. A. 2006b. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell, 125***,** *301-313.*

*LEE, W. H., BOOKSTEIN, R., HONG, F., YOUNG, L. J., SHEW, J. Y. & LEE, E. Y. 1987. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. Science, 235***,** *1394-9.*

*LEE, Y., AHN, C., HAN, J., CHOI, H., KIM, J., YIM, J., LEE, J., PROVOST, P., RADMARK, O., KIM, S. & KIM, V. N. 2003b. The nuclear RNase III Drosha initiates microRNA processing. Nature, 425***,** *415-9.*

*LEE, Y., JEON, K., LEE, J. T., KIM, S. & KIM, V. N. 2002. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J, 21***,** *4663-70.*

*LENA, A. M., SHALOM-FEUERSTEIN, R., RIVETTI DI VAL CERVO, P., ABERDAM, D., KNIGHT, R. A., MELINO, G. & CANDI, E. 2008. miR-203 represses 'stemness' by repressing DeltaNp63. Cell Death Differ, 15***,** *1187-95.*

LENTINI, M., SCHEPIS, C., CUPPARI, D. A. & BATOLO, D. 2006. Tenascin expression in actinic keratosis. J Cutan Pathol, 33, *716-20.*

LERMAN, G., AVIVI, C., MARDOUKH, C., BARZILAI, A., TESSONE, A., GRADUS, B., PAVLOTSKY, F., BARSHACK, I., POLAK-CHARCON, S., ORENSTEIN, A., HORNSTEIN, E., SIDI, Y. & AVNI, D. 2011. MiRNA expression in psoriatic skin: reciprocal regulation of hsa-miR-99a and IGF-1R. PloS one, 6, *e20916.*

LEVY, C., KHALED, M., ILIOPOULOS, D., JANAS, M. M., SCHUBERT, S., PINNER, S., CHEN, P. H., LI, S., FLETCHER, A. L., YOKOYAMA, S., SCOTT, K. L., GARRAWAY, L. A., SONG, J. S., GRANTER, S. R., TURLEY, S. J., FISHER, D. E. & NOVINA, C. D. 2010. Intronic miR-211 assumes the tumor suppressive function of its host gene in melanoma. Molecular cell, 40, *841-9.*

LEWIS, B. P., BURGE, C. B. & BARTEL, D. P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell, 120, *15-20.*

LI, F., MAO, G., TONG, D., HUANG, J., GU, L., YANG, W. & LI, G.-M. 2013a. The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through Its Interaction with MutSα. Cell, 153, *590-600.*

LI, G.-M. 2013. Decoding the Histone Code: Role of H3K36me3 in Mismatch Repair and Implications for Cancer Susceptibility and Therapy. Cancer Research, 73, *6379-6383.*

LI, J., WANG, K., GAO, F., JENSEN, T. D., LI, S. T., DEANGELIS, P. M., KOLVRAA, S., PROBY, C., FORSLUND, O., BOLUND, L. & CLAUSEN, O. P. 2012. Array comparative genomic hybridization of keratoacanthomas and squamous cell carcinomas: different patterns of genetic aberrations suggest two distinct entities. J Invest Dermatol, 132, *2060-6.*

LI, Q., YANG, X. H., XU, F., SHARMA, C., WANG, H. X., KNOBLICH, K., RABINOVITZ, I., GRANTER, S. R. & HEMLER, M. E. 2013b. Tetraspanin CD151 plays a key role in skin squamous cell carcinoma. Oncogene, 32, *1772-83.*

LIAN, C. G., XU, Y., CEOL, C., WU, F., LARSON, A., DRESSER, K., XU, W., TAN, L., HU, Y., ZHAN, Q., LEE, C. W., HU, D., LIAN, B. Q., KLEFFEL, S., YANG, Y., NEISWENDER, J., KHORASANI, A. J., FANG, R., LEZCANO, C., DUNCAN, L. M., SCOLYER, R. A., THOMPSON, J. F., KAKAVAND, H., HOUVRAS, Y., ZON, L. I., MIHM, M. C., JR., KAISER, U. B., SCHATTON, T., WODA, B. A., MURPHY, G. F. & SHI, Y. G. 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell, 150, *1135-46.*

LIBERZON, A., SUBRAMANIAN, A., PINCHBACK, R., THORVALDSDÓTTIR, H., TAMAYO, P. & MESIROV, J. P. 2011. Molecular signatures database (MSigDB) 3.0. Bioinformatics, 27, *1739-1740.*

LIND, G. E., SKOTHEIM, R. I., FRAGA, M. F., ABELER, V. M., ESTELLER, M. & LOTHE, R. A. 2006. Novel epigenetically deregulated genes in testicular cancer include homeobox genes and SCGB3A1 (HIN-1). The Journal of pathology, 210, *441-9.*

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature, 462, *315-22.*

LISTER, R., PELIZZOLA, M., KIDA, Y. S., HAWKINS, R. D., NERY, J. R., HON, G., ANTOSIEWICZ-BOURGET, J., O'MALLEY, R., CASTANON, R., KLUGMAN, S., DOWNES, M., YU, R., STEWART, R., REN, B., THOMSON, J. A., EVANS, R. M. & ECKER, J. R. 2011. Hotspots of *aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature, 471, 68-73.*

LIU, Y. & WU, F. 2010. Global burden of aflatoxin-induced hepatocellular carcinoma: a risk assessment. Environ Health Perspect, 118, 818-24.

LJUJIC, B., MILOVANOVIC, M., VOLAREVIC, V., MURRAY, B., BUGARSKI, D., PRZYBORSKI, S., ARSENIJEVIC, N., LUKIC, M. L. & STOJKOVIC, M. 2013. Human mesenchymal stem cells creating an immunosuppressive environment and promote breast cancer in mice. Sci. Rep., 3.

LODYGIN, D., YAZDI, A. S., SANDER, C. A., HERZINGER, T. & HERMEKING, H. 2003. Analysis of 14-3-3sigma expression in hyperproliferative skin diseases reveals selective loss associated with CpG-methylation in basal cell carcinoma. Oncogene, 22, 5519-24.

LONGWORTH, M. S. & LAIMINS, L. A. 2004. Pathogenesis of Human Papillomaviruses in Differentiating Epithelia. Microbiology and Molecular Biology Reviews, 68, 362-372.

LÓPEZ-BIGAS, N., MELCHIONDA, S., GASPARINI, P., BORRAGÁN, A., ARBONÉS, M. L. & ESTIVILL, X. 2002. A common frameshift mutation and other variants in GJB4 (connexin 30.3): Analysis of hearing impairment families. Human Mutation, 19, 458-458.

LORSBACH, R. B., MOORE, J., MATHEW, S., RAIMONDI, S. C., MUKATIRA, S. T. & DOWNING, J. R. 2003. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia, 17, 637-41.

LOWE, R., GEMMA, C., BEYAN, H., HAWA, M. I., BAZEOS, A., LESLIE, R. D., MONTPETIT, A., RAKYAN, V. K. & RAMAGOPALAN, S. V. 2013. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. Epigenetics, 8.

LU, J., GETZ, G., MISKA, E. A., ALVAREZ-SAAVEDRA, E., LAMB, J., PECK, D., SWEET-CORDERO, A., EBERT, B. L., MAK, R. H., FERRANDO, A. A., DOWNING, J. R., JACKS, T., HORVITZ, H. R. & GOLUB, T. R. 2005. MicroRNA expression profiles classify human cancers. Nature, 435, 834-8.

MADAN, V., LEAR, J. T. & SZEIMIES, R. M. 2010. Non-melanoma skin cancer. Lancet, 375, 673-85.

MADISON, B. B., MCKENNA, L. B., DOLSON, D., EPSTEIN, D. J. & KAESTNER, K. H. 2009. FoxF1 and FoxL1 link hedgehog signaling and the control of epithelial proliferation in the developing stomach and intestine. J Biol Chem, 284, 5936-44.

MAGEE, JEFFREY A., PISKOUNOVA, E. & MORRISON, SEAN J. 2012. Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. Cancer Cell, 21, 283-296.

MAITRA, A., ARKING, D. E., SHIVAPURKAR, N., IKEDA, M., STASTNY, V., KASSAUEI, K., SUI, G., CUTLER, D. J., LIU, Y., BRIMBLE, S. N., NOAKSSON, K., HYLLNER, J., SCHULZ, T. C., ZENG, X., FREED, W. J., CROOK, J., ABRAHAM, S., COLMAN, A., SARTIPY, P., MATSUI, S., CARPENTER, M., GAZDAR, A. F., RAO, M. & CHAKRAVARTI, A. 2005. Genomic alterations in cultured human embryonic stem cells. Nat Genet, 37, 1099-103.

MAKSIMOVIC, J., GORDON, L. & OSHLACK, A. 2012. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol, 13, R44.

MALTEPE, E., BAKARDJIEV, A. I. & FISHER, S. J. 2010. The placenta: transcriptional, epigenetic, and physiological integration during development. J Clin Invest, 120, 1016-25.

MANNA, S., SINGHA, B., PHYO, S. A., GATLA, H. R., CHANG, T. P., SANACORA, S., RAMASWAMI, S. & VANCUROVA, I. 2013. Proteasome inhibition by bortezomib increases IL-8 expression in *androgen-independent prostate cancer cells: the role of IKKalpha. J Immunol, 191***,** *2837-46.*

*MARCHINI, S., MARRAZZO, E., BONOMI, R., CHIORINO, G., ZAFFARONI, M., WEISSBACH, L., HORNICEK, F. J., BROGGINI, M., FAIRCLOTH, G. T. & D'INCALCI, M. 2005. Molecular characterisation of two human cancer cell lines selected in vitro for their chemotherapeutic drug resistance to ET-743. Eur J Cancer, 41***,** *323-33.*

*MARIN, R. M., SULC, M. & VANICEK, J. 2013. Searching the coding region for microRNA targets. RNA, 19***,** *467-74.*

*MARQUEZ, J., KOHLI, M., ARTETA, B., CHANG, S., LI, W. B., GOLDBLATT, M. & VIDAL-VANACLOCHA, F. 2013. Identification of hepatic microvascular adhesion-related genes of human colon cancer cells using random homozygous gene perturbation. Int J Cancer, 133***,** *2113-22.*

*MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proceedings of the National Academy of Sciences, 78***,** *7634-7638.*

*MARTINEZ-CRUZ, A. B., SANTOS, M., LARA, M. F., SEGRELLES, C., RUIZ, S., MORAL, M., LORZ, C., GARCIA-ESCUDERO, R. & PARAMIO, J. M. 2008. Spontaneous squamous cell carcinoma induced by the somatic inactivation of retinoblastoma and Trp53 tumor suppressors. Cancer Res, 68***,** *683-92.*

*MASRI, S. & SASSONE-CORSI, P. 2010. Plasticity and specificity of the circadian epigenome. Nat Neurosci, 13***,** *1324-9.*

*MATHSYARAJA, H. & OSTROWSKI, M. C. 2012. Setting Snail2's pace during EMT. Nat Cell Biol, 14***,** *1122-3.*

*MATSUURA, M., SUZUKI, T. & SAITO, T. 2010. Osteopontin is a new target molecule for ovarian clear cell carcinoma therapy. Cancer Sci, 101***,** *1828-33.*

*MAUNAKEA, A. K., NAGARAJAN, R. P., BILENKY, M., BALLINGER, T. J., D'SOUZA, C., FOUSE, S. D., JOHNSON, B. E., HONG, C., NIELSEN, C., ZHAO, Y., TURECKI, G., DELANEY, A., VARHOL, R., THIESSEN, N., SHCHORS, K., HEINE, V. M., ROWITCH, D. H., XING, X., FIORE, C., SCHILLEBEECKX, M., JONES, S. J., HAUSSLER, D., MARRA, M. A., HIRST, M., WANG, T. & COSTELLO, J. F. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature, 466***,** *253-7.*

*MAURER, B., STANCZYK, J., JUNGEL, A., AKHMETSHINA, A., TRENKMANN, M., BROCK, M., KOWAL-BIELECKA, O., GAY, R. E., MICHEL, B. A., DISTLER, J. H., GAY, S. & DISTLER, O. 2010. MicroRNA-29, a key regulator of collagen expression in systemic sclerosis. Arthritis Rheum, 62***,** *1733-43.*

*MCGREGOR, J. M. & PROBY, C. M. 1996. The role of papillomaviruses in human non-melanoma skin cancer. Cancer Surv, 26***,** *219-36.*

MCKINNELL, R. G. 1998. The biological basis of cancer, Cambridge, Cambridge University Press.

*MELINO, G. 2011. p63 is a suppressor of tumorigenesis and metastasis interacting with mutant p53. Cell Death Differ, 18***,** *1487-1499.*

*MENDEZ, E., FAN, W., CHOI, P., AGOFF, S. N., WHIPPLE, M., FARWELL, D. G., FUTRAN, N. D., WEYMULLER, E. A., JR., ZHAO, L. P. & CHEN, C. 2007. Tumor-specific genetic expression profile of metastatic oral squamous cell carcinoma. Head Neck, 29***,** *803-14.*

MENDEZ-FERRER, S., MICHURINA, T. V., FERRARO, F., MAZLOOM, A. R., MACARTHUR, B. D., LIRA, S. A., SCADDEN, D. T., MA'AYAN, A., ENIKOLOPOV, G. N. & FRENETTE, P. S. 2010.

*Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. Nature, 466***,** *829-34.*

*MERCHANT, A. A. & MATSUI, W. 2010. Targeting Hedgehog--a cancer stem cell pathway. Clin Cancer Res, 16***,** *3130-40.*

*MERTZ, K. D., PFALTZ, M., JUNT, T., SCHMID, M., FERNANDEZ FIGUERAS, M. T., PFALTZ, K., BARGHORN, A. & KEMPF, W. 2010. Merkel cell polyomavirus is present in common warts and carcinoma in situ of the skin. Hum Pathol, 41***,** *1369-79.*

*MESHORER, E. & MISTELI, T. 2006. Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol, 7***,** *540-546.*

*MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T.-K., KOCHE, R. P., LEE, W., MENDENHALL, E., O/'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature, 448***,** *553-560.*

*MILLAR, S. E. 2011. Committing to a hairy fate: epigenetic regulation of hair follicle stem cells. Cell Stem Cell, 9***,** *183-4.*

*MILLER, R. W. & RABKIN, C. S. 1999. Merkel cell carcinoma and melanoma: etiological similarities and differences. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 8***,** *153-8.*

*MILLS, A. A., ZHENG, B., WANG, X. J., VOGEL, H., ROOP, D. R. & BRADLEY, A. 1999. p63 is a p53 homologue required for limb and epidermal morphogenesis. Nature, 398***,** *708-13.*

*MISAWA, K., UEDA, Y., KANAZAWA, T., MISAWA, Y., JANG, I., BRENNER, J. C., OGAWA, T., TAKEBAYASHI, S., GRENMAN, R. A., HERMAN, J. G., MINETA, H. & CAREY, T. E. 2008. Epigenetic inactivation of galanin receptor 1 in head and neck cancer. Clin Cancer Res, 14***,** *7604-13.*

*MISHRA, P., HUMENIUK, R., MEDINA, D., ALEXE, G., MESIROV, J., GANESAN, S., GLOD, J. & BANERJEE, D. 2008. Carcinoma-associated fibroblast-like differentiation of human mesenchymal stem cells. Cancer Res, 68***,** *4331 - 9.*

*MIYAZONO, K. 2009. Transforming growth factor-beta signaling in epithelial-mesenchymal transition and progression of cancer. Proc Jpn Acad Ser B Phys Biol Sci, 85***,** *314-23.*

*MIYOSHI, H., SHIMIZU, K., KOZU, T., MASEKI, N., KANEKO, Y. & OHKI, M. 1991. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. Proc Natl Acad Sci U S A, 88***,** *10431-4.*

*MLADKOVA, N. & CHAKRAVARTI, A. 2009. Molecular profiling in glioblastoma: prelude to personalized treatment. Curr Oncol Rep, 11***,** *53-61.*

*MOGHA, A., FAUTREL, A., MOUCHET, N., GUO, N., CORRE, S., ADAMSKI, H., WATIER, E., MISERY, L. & GALIBERT, M. D. 2010. Merkel cell polyomavirus small T antigen mRNA level is increased following in vivo UV-radiation. PloS one, 5***,** *e11423.*

*MOHAMED, A. A., TAN, S. H., SUN, C., SHAHEDUZZAMAN, S., HU, Y., PETROVICS, G., CHEN, Y., SESTERHENN, I. A., LI, H., SREENATH, T., MCLEOD, D. G., DOBI, A. & SRIVASTAVA, S. 2011. ERG oncogene modulates prostaglandin signaling in prostate cancer cells. Cancer Biol Ther, 11***,** *410-7.*

MOORE, P. S. & CHANG, Y. 1995. Detection of Herpesvirus-Like DNA Sequences in Kaposi's Sarcoma in Patients with and Those without HIV Infection. New England Journal of Medicine, 332, ***,*** *1181-1185.*

MORAN-CRUSIO, K., REAVIE, L., SHIH, A., ABDEL-WAHAB, O., NDIAYE-LOBRY, D., LOBRY, C., FIGUEROA, M. E., VASANTHAKUMAR, A., PATEL, J., ZHAO, X., PERNA, F., PANDEY, S., MADZO, J., SONG, C., DAI, Q., HE, C., IBRAHIM, S., BERAN, M., ZAVADIL, J., NIMER, S. D., MELNICK, A., GODLEY, L. A., AIFANTIS, I. & LEVINE, R. L. 2011. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell, 20, *11-24.*

MORRIS, T. & LOWE, R. 2012. Report on the Infinium 450k methylation array analysis workshop: April 20, 2012 UCL, London, UK. Epigenetics, 7, *961-2.*

MORSE, D., DAILEY, R. C. & BUNN, J. 1974. Prehistoric multiple myeloma. Bull N Y Acad Med, 50, *447-58.*

MOTIWALA, T. & JACOB, S. T. 2006. Role of protein tyrosine phosphatases in cancer. Prog Nucleic Acid Res Mol Biol, 81, *297-329.*

MOUBAYED, N., WEICHENTHAL, M., HARDER, J., WANDEL, E., STICHERLING, M. & GLASER, R. 2007. Psoriasin (S100A7) is significantly up-regulated in human epithelial skin tumours. J Cancer Res Clin Oncol, 133, *253-61.*

MULDER, K. W., WANG, X., ESCRIU, C., ITO, Y., SCHWARZ, R. F., GILLIS, J., SIROKMANY, G., DONATI, G., URIBE-LEWIS, S., PAVLIDIS, P., MURRELL, A., MARKOWETZ, F. & WATT, F. M. 2012. Diverse epigenetic strategies interact to control epidermal differentiation. Nat Cell Biol, 14, *753-763.*

MURAKAMI, M., IMAJOH, M., IKAWA, T., NAKAJIMA, H., KAMIOKA, M., NEMOTO, Y., UJIHARA, T., UCHIYAMA, J., MATSUZAKI, S., SANO, S. & DAIBATA, M. 2011. Presence of Merkel cell polyomavirus in Japanese cutaneous squamous cell carcinoma. Journal of Clinical Virology, 50, *37-41.*

MURAO, K., KUBO, Y., OHTANI, N., HARA, E. & ARASE, S. 2006. Epigenetic abnormalities in cutaneous squamous cell carcinomas: frequent inactivation of the RB1/p16 and p53 pathways. The British journal of dermatology, 155, *999-1005.*

MURPHY, G. M., NORRIS, P. G., YOUNG, A. R., CORBETT, M. F. & HAWK, J. L. 1993. Low-dose ultraviolet-B irradiation depletes human epidermal Langerhans cells. Br J Dermatol, 129, *674-7.*

MURRY, C. E. & KELLER, G. 2008. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell, 132, *661-80.*

NAIR, M., TENG, A., BILANCHONE, V., AGRAWAL, A., LI, B. & DAI, X. 2006. Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription. J Cell Biol, 173, *253-64.*

NAKABAYASHI, K., TRUJILLO, A. M., TAYAMA, C., CAMPRUBI, C., YOSHIDA, W., LAPUNZINA, P., SANCHEZ, A., SOEJIMA, H., ABURATANI, H., NAGAE, G., OGATA, T., HATA, K. & MONK, D. 2011. Methylation screening of reciprocal genome-wide UPDs identifies novel human-specific imprinted genes. Hum Mol Genet, 20, *3188-97.*

NAKAMURA, M. & TOKURA, Y. 2011. Epithelial-mesenchymal transition in the skin. J Dermatol Sci, 61, *7-13.*

*NALBANT, D., YOUN, H., NALBANT, S. I., SHARMA, S., COBOS, E., BEALE, E. G., DU, Y. & WILLIAMS, S. C. 2005. FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. BMC Genomics, 6***,** *11.*

*NANTY, L., CARBAJOSA, G., HEAP, G. A., RATNIEKS, F., VAN HEEL, D. A., DOWN, T. A. & RAKYAN, V. K. 2011. Comparative methylomics reveals gene-body H3K36me3 in Drosophila predicts DNA methylation and CpG landscapes in other invertebrates. Genome Res, 21***,** *1841-50.*

*NATARAJ, A. J., BLACK, H. S. & ANANTHASWAMY, H. N. 1996. Signature p53 mutation at DNA cross-linking sites in 8-methoxypsoralen and ultraviolet A (PUVA)-induced murine skin cancers. Proc Natl Acad Sci U S A, 93***,** *7961-5.*

*NAVIN, N., KENDALL, J., TROGE, J., ANDREWS, P., RODGERS, L., MCINDOO, J., COOK, K., STEPANSKY, A., LEVY, D., ESPOSITO, D., MUTHUSWAMY, L., KRASNITZ, A., MCCOMBIE, W. R., HICKS, J. & WIGLER, M. 2011. Tumour evolution inferred by single-cell sequencing. Nature, 472***,** *90-4.*

NEAL, A. J. & HOSKIN, P. J. 2009. Clinical oncology : basic principles and practice, London, Hodder Arnold.

*NEALE, R. E., WEISSENBORN, S., ABENI, D., BAVINCK, J. N. B., EUVRARD, S., FELTKAMP, M. C. W., GREEN, A. C., HARWOOD, C., DE KONING, M., NALDI, L., NINDL, I., PAWLITA, M., PROBY, C., QUINT, W. G., WATERBOER, T., WIELAND, U. & PFISTER, H. 2013. Human Papillomavirus Load in Eyebrow Hair Follicles and Risk of Cutaneous Squamous Cell Carcinoma. Cancer Epidemiology Biomarkers & Prevention, 22***,** *719-727.*

*NERLICH, A. G., ROHRBACH, H., BACHMEIER, B. & ZINK, A. 2006. Malignant tumors in two ancient populations: An approach to historical tumor epidemiology. Oncol Rep, 16***,** *197-202.*

*NGUYEN, H., MERRILL, B. J., POLAK, L., NIKOLOVA, M., RENDL, M., SHAVER, T. M., PASOLLI, H. A. & FUCHS, E. 2009. Tcf3 and Tcf4 are essential for long-term homeostasis of skin epithelia. Nat Genet, 41***,** *1068-75.*

*NILMANAT, K., CHAILUNGKA, P., TEMSAK, P., PROMNOI, C., TULATHAMKIT, K., NOO-URAI, P. & PHATTARANAVIG, S. 2010. Living with suffering as voiced by Thai patients with terminal advanced cancer. Int J Palliat Nurs, 16***,** *393-9.*

*NINDL, I., DANG, C., FORSCHNER, T., KUBAN, R. J., MEYER, T., STERRY, W. & STOCKFLETH, E. 2006. Identification of differentially expressed genes in cutaneous squamous cell carcinoma by microarray expression profiling. Mol Cancer, 5***,** *30.*

*NING, H., YANG, F., JIANG, M., HU, L., FENG, K., ZHANG, J., YU, Z., LI, B., XU, C., LI, Y., WANG, J., HU, J., LOU, X. & CHEN, H. 2008. The correlation between cotransplantation of mesenchymal stem cells and higher recurrence rate in hematologic malignancy patients: outcome of a pilot clinical study. Leukemia, 22***,** *593 - 9.*

*NOGGLE, S., FUNG, H. L., GORE, A., MARTINEZ, H., SATRIANI, K. C., PROSSER, R., OUM, K., PAULL, D., DRUCKENMILLER, S., FREEBY, M., GREENBERG, E., ZHANG, K., GOLAND, R., SAUER, M. V., LEIBEL, R. L. & EGLI, D. 2011. Human oocytes reprogram somatic cells to a pluripotent state. Nature, 478***,** *70-5.*

*NOTARI, M., HU, Y., KOCH, S., LU, M., RATNAYAKA, I., ZHONG, S., BAER, C., PAGOTTO, A., GOLDIN, R., SALTER, V., CANDI, E., MELINO, G. & LU, X. 2011. Inhibitor of apoptosis-stimulating protein of p53 (iASPP) prevents senescence and is required for epithelial stratification. Proceedings of the National Academy of Sciences of the United States of America, 108***,** *16645-50.*

*NOWELL, P. C. 1976. The clonal evolution of tumor cell populations. Science, 194***,** *23-8.*

*NUNES, F. D., DE ALMEIDA, F. C., TUCCI, R. & DE SOUSA, S. C. 2003. Homeobox genes: a molecular link between development and cancer. Pesqui Odontol Bras, 17***,** *94-8.*

*NURUL-SYAKIMA, A. M., YOKE-KQUEEN, C., SABARIAH, A. R., SHIRAN, M. S., SINGH, A. & LEARN-HAN, L. 2011. Differential microRNA expression and identification of putative miRNA targets and pathways in head and neck cancers. Int J Mol Med, 28***,** *327-36.*

*O'BRYAN, K., SHERMAN, W., NIEDT, G. W., TABACK, B., MANOLIDIS, S., WANG, A. & RATNER, D. 2013. An evolving paradigm for the workup and management of high-risk cutaneous squamous cell carcinoma. J Am Acad Dermatol, 69***,** *595-602 e1.*

*O'CONNOR, P. M., JACKMAN, J., BAE, I., MYERS, T. G., FAN, S., MUTOH, M., SCUDIERO, D. A., MONKS, A., SAUSVILLE, E. A., WEINSTEIN, J. N., FRIEND, S., FORNACE, A. J., JR. & KOHN, K. W. 1997. Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. Cancer Res, 57***,** *4285-300.*

*OBERHOLZER, P. A., KEE, D., DZIUNYCZ, P., SUCKER, A., KAMSUKOM, N., JONES, R., RODEN, C., CHALK, C. J., ARDLIE, K., PALESCANDOLO, E., PIRIS, A., MACCONAILL, L. E., ROBERT, C., HOFBAUER, G. F., MCARTHUR, G. A., SCHADENDORF, D. & GARRAWAY, L. A. 2012. RAS mutations are associated with the development of cutaneous squamous cell tumors in patients treated with RAF inhibitors. J Clin Oncol, 30***,** *316-21.*

*OBERST, A., MALATESTA, M., AQEILAN, R. I., ROSSI, M., SALOMONI, P., MURILLAS, R., SHARMA, P., KUEHN, M. R., OREN, M., CROCE, C. M., BERNASSOLA, F. & MELINO, G. 2007. The Nedd4-binding partner 1 (N4BP1) protein is an inhibitor of the E3 ligase Itch. Proc Natl Acad Sci U S A, 104***,** *11280-5.*

*OHM, J. E., MCGARVEY, K. M., YU, X., CHENG, L., SCHUEBEL, K. E., COPE, L., MOHAMMAD, H. P., CHEN, W., DANIEL, V. C., YU, W., BERMAN, D. M., JENUWEIN, T., PRUITT, K., SHARKIS, S. J., WATKINS, D. N., HERMAN, J. G. & BAYLIN, S. B. 2007. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet, 39***,** *237-42.*

*OHYAMA, M. 2007. Hair follicle bulge: a fascinating reservoir of epithelial stem cells. Journal of dermatological science, 46***,** *81-9.*

*OHYAMA, M., TERUNUMA, A., TOCK, C. L., RADONOVICH, M. F., PISE-MASISON, C. A., HOPPING, S. B., BRADY, J. N., UDEY, M. C. & VOGEL, J. C. 2006. Characterization and isolation of stem cell-enriched human hair follicle bulge cells. J Clin Invest, 116***,** *249-60.*

*ONDA, T., UZAWA, K., ENDO, Y., BUKAWA, H., YOKOE, H., SHIBAHARA, T. & TANZAWA, H. 2006. Ubiquitous mitochondrial creatine kinase downregulated in oral squamous cell carcinoma. British journal of cancer, 94***,** *698-709.*

*ORMESTAD, M., ASTORGA, J., LANDGREN, H., WANG, T., JOHANSSON, B. R., MIURA, N. & CARLSSON, P. 2006. Foxf1 and Foxf2 control murine gut development by limiting mesenchymal Wnt signaling and promoting extracellular matrix production. Development, 133***,** *833-43.*

*ORTH, G. 1986. Epidermodysplasia verruciformis: a model for understanding the oncogenicity of human papillomaviruses. Ciba Found Symp, 120***,** *157-74.*

*ORTH, G., FAVRE, M. & CROISSANT, O. 1977. Characterization of a New Type of Human Papillomavirus That Causes Skin Warts. J. Virol., 24***,** *108-120.*

*ORTNER, D. J., PONCE, P., OGDEN, A. & BUCKBERRY, J. 2010. Multicentric osteosarcoma associated with DISH, in a 19th century burial from England. International Journal of Osteoarchaeology***,** *n/a-n/a.*

OZASA, K., SHIMIZU, Y., SAKATA, R., SUGIYAMA, H., GRANT, E. J., SODA, M., KASAGI, F. & SUYAMA, A. 2011. Risk of Cancer and Non-Cancer Diseases in the Atomic Bomb Survivors. Radiat Prot Dosimetry.

PADMANABHAN, A., LI, X. & BIEBERICH, C. J. 2013. Protein kinase A regulates MYC protein through transcriptional and post-translational mechanisms in a catalytic subunit isoform-specific manner. J Biol Chem, 288, *14158-69.*

PARK, J., BRENA, R. M., GRUIDL, M., ZHOU, J., HUANG, T., PLASS, C. & TOCKMAN, M. S. 2005. CpG island hypermethylation profiling of lung cancer using restriction landmark genomic scanning (RLGS) analysis. Cancer biomarkers : section A of Disease markers, 1, *193-200.*

PARKIN, D. M. 2006. The global health burden of infection-associated cancers in the year 2002. International Journal of Cancer, 118, *3030-3044.*

PARSONNET, J. 1999. Microbes and malignancy : infection as a cause of human cancers, New York, Oxford University Press.

PATEL, A. S., KARAGAS, M. R., PAWLITA, M., WATERBOER, T. & NELSON, H. H. 2008. Cutaneous human papillomavirus infection, the EVER2 gene and incidence of squamous cell carcinoma: a case-control study. Int J Cancer, 122, *2377-9.*

PATEL, G. K., YEE, C. L., TERUNUMA, A., TELFORD, W. G., VOONG, N., YUSPA, S. H. & VOGEL, J. C. 2012. Identification and characterization of tumor-initiating cells in human primary cutaneous squamous cell carcinoma. J Invest Dermatol, 132, *401-9.*

PATEL, V., ROSENFELDT, H. M., LYONS, R., SERVITJA, J. M., BUSTELO, X. R., SIROFF, M. & GUTKIND, J. S. 2007. Persistent activation of Rac1 in squamous carcinomas of the head and neck: evidence for an EGFR/Vav2 signaling axis involved in cell invasion. Carcinogenesis, 28, *1145-52.*

PAULSON, K. G., CARTER, J. J., JOHNSON, L. G., CAHILL, K. W., IYER, J. G., SCHRAMA, D., BECKER, J. C., MADELEINE, M. M., NGHIEM, P. & GALLOWAY, D. A. 2010. Antibodies to merkel cell polyomavirus T antigen oncoproteins reflect tumor burden in merkel cell carcinoma patients. Cancer research, 70, *8388-97.*

PAZ, M. F., FRAGA, M. F., AVILA, S., GUO, M., POLLAN, M., HERMAN, J. G. & ESTELLER, M. 2003. A Systematic Profile of DNA Methylation in Human Cancer Cell Lines. Cancer Research, 63, *1114-1121.*

PAZDUR, R. 2010. FDA Approval for Bevacizumab [Online]. National Cancer Institute. Available: http://www.cancer.gov/cancertopics/druginfo/fda-bevacizumab *[Accessed 19.5.2011 2011].*

PELISSON, I., SOLER, C., CHARDONNET, Y., EUVRARD, S. & SCHMITT, D. 1996. A Possible Role for Human Papillomaviruses and c-myc, c-Ha-ras, And p53 Gene Alterations in Malignant Cutaneous Lesions From Renal Transplant Recipients. Cancer Detection and Prevention, 20, *20-30.*

PERA, M. F. & TROUNSON, A. O. 2004. Human embryonic stem cells: prospects for development. Development, 131, *5515-25.*

PEREZ, C. A., OTT, J., MAYS, D. J. & PIETENPOL, J. A. 2007. p63 consensus DNA-binding site: identification, analysis and application into a p63MH algorithm. Oncogene, 26, *7363-70.*

PETO, J., MATTHEWS, F. E., HODGSON, J. T. & JONES, J. R. 1995. Continuing increase in mesothelioma mortality in Britain. The Lancet, 345, *535-539.*

PETTER, G. & HAUSTEIN, U.-F. 2000. Histologic Subtyping and Malignancy Assessment of Cutaneous Squamous Cell Carcinoma. Dermatologic Surgery, 26, *521-530.*

*PFEFFER, S. & VOINNET, O. 2006. Viruses, microRNAs and cancer. Oncogene, 25***,** *6211-9.*

*PHAN, T. A., HALLIDAY, G. M., BARNETSON, R. S. & DAMIAN, D. L. 2006. Spectral and dose dependence of ultraviolet radiation-induced immunosuppression. Front Biosci, 11***,** *394-411.*

*PICCIRILLO, S. G., COMBI, R., CAJOLA, L., PATRIZI, A., REDAELLI, S., BENTIVEGNA, A., BARONCHELLI, S., MAIRA, G., POLLO, B., MANGIOLA, A., DIMECO, F., DALPRA, L. & VESCOVI, A. L. 2009. Distinct pools of cancer stem-like cells coexist within human glioblastomas and display different tumorigenicity and independent genomic evolution. Oncogene, 28***,** *1807-11.*

*PIDSLEY, R., CC, Y. W., VOLTA, M., LUNNON, K., MILL, J. & SCHALKWYK, L. C. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics, 14***,** *293.*

*PIEKARZ, R. L. & BATES, S. E. 2009. Epigenetic modifiers: basic understanding and clinical development. Clin Cancer Res, 15***,** *3918-26.*

*PIETERSEN, A. M. & VAN LOHUIZEN, M. 2008. Stem cell regulation by polycomb repressors: postponing commitment. Current Opinion in Cell Biology, 20***,** *201-207.*

*PILCH, H., WEISS, J., HEUBNER, C. & HEINE, M. 1994. Differential diagnosis of keratoacanthomas and squamous cell carcinomas: diagnostic value of DNA image cytometry and p53 expression. J Cutan Pathol, 21***,** *507-13.*

*PITTENGER, M. F., MACKAY, A. M., BECK, S. C., JAISWAL, R. K., DOUGLAS, R., MOSCA, J. D., MOORMAN, M. A., SIMONETTI, D. W., CRAIG, S. & MARSHAK, D. R. 1999. Multilineage potential of adult human mesenchymal stem cells. Science, 284***,** *143-7.*

*PIYATHILAKE, C. J., HENAO, O., FROST, A. R., MACALUSO, M., BELL, W. C., JOHANNING, G. L., HEIMBURGER, D. C., NIVELEAU, A. & GRIZZLE, W. E. 2003. Race- and age-dependent alterations in global methylation of DNA in squamous cell carcinoma of the lung (United States). Cancer causes & control : CCC, 14***,** *37-42.*

*POCHAMPALLY, R. R., YLOSTALO, J., PENFORNIS, P., MATZ, R. R., SMITH, J. R. & PROCKOP, D. J. 2007. Histamine receptor H1 and dermatopontin: new downstream targets of the vitamin D receptor. J Bone Miner Res, 22***,** *1338-49.*

*POLYAK, K., HAVIV, I. & CAMPBELL, I. G. 2009. Co-evolution of tumor cells and their microenvironment. Trends in Genetics, 25***,** *30-38.*

*POTTER, C. S., PRUETT, N. D., KERN, M. J., BAYBO, M. A., GODWIN, A. R., POTTER, K. A., PETERSON, R. L., SUNDBERG, J. P. & AWGULEWITSCH, A. 2011. The nude mutant gene Foxn1 is a HOXC13 regulatory target during hair follicle and nail differentiation. J Invest Dermatol, 131***,** *828-37.*

*POULIN, D. L. & DECAPRIO, J. A. 2006. Is there a role for SV40 in human cancer? J Clin Oncol, 24***,** *4356-65.*

*POURREYRON, C., COX, G., MAO, X., VOLZ, A., BAKSH, N., WONG, T., FASSIHI, H., ARITA, K., O'TOOLE, E. A., OCAMPO-CANDIANI, J., CHEN, M., HART, I. R., BRUCKNER-TUDERMAN, L., SALAS-ALANIS, J. C., MCGRATH, J. A., LEIGH, I. M. & SOUTH, A. P. 2007. Patients with recessive dystrophic epidermolysis bullosa develop squamous-cell carcinoma regardless of type VII collagen expression. J Invest Dermatol, 127***,** *2438-44.*

PROBY, C. M., HARWOOD, C. A., NEALE, R. E., GREEN, A. C., EUVRARD, S., NALDI, L., TESSARI, G., FELTKAMP, M. C. W., DE KONING, M. N. C., QUINT, W. G. V., WATERBOER, T., PAWLITA, M., WEISSENBORN, S., WIELAND, U., PFISTER, H., STOCKFLETH, E., NINDL, I., ABENI, D.,

*SCHEGGET, J. T., BOUWES BAVINCK, J. N. & THE, E. P. I. H. P. V. U. V. C. A. G. 2011. A Case–Control Study of Betapapillomavirus Infection and Cutaneous Squamous Cell Carcinoma in Organ Transplant Recipients. American Journal of Transplantation, 11***,** *1498-1508.*

*PURDIE, K. J., HARWOOD, C. A., GULATI, A., CHAPLIN, T., LAMBERT, S. R., CERIO, R., KELLY, G. P., CAZIER, J.-B., YOUNG, B. D., LEIGH, I. M. & PROBY, C. M. 2009. Single Nucleotide Polymorphism Array Analysis Defines a Specific Genetic Fingerprint for Well-Differentiated Cutaneous SCCs. J Invest Dermatol, 129***,** *1562-1568.*

*QIAO, L., XU, Z., ZHAO, T., YE, L. & ZHANG, X. 2008a. Dkk-1 secreted by mesenchymal stem cells inhibits growth of breast cancer cells via depression of Wnt signalling. Cancer Lett, 269***,** *67 - 77.*

*QIAO, L., XU, Z., ZHAO, T., ZHAO, Z., SHI, M., ZHAO, R., YE, L. & ZHANG, X. 2008b. Suppression of tumorigenesis by human mesenchymal stem cells in a hepatoma model. Cell Res, 18***,** *500 - 7.*

*QUAEDVLIEG, P. J., TIRSI, E., THISSEN, M. R. & KREKELS, G. A. 2006. Actinic keratosis: how to differentiate the good from the bad ones? Eur J Dermatol, 16***,** *335-9.*

*QUINN, A. G., SIKKINK, S. & REES, J. L. 1994. Basal Cell Carcinomas and Squamous Cell Carcinomas of Human Skin Show Distinct Patterns of Chromosome Loss. Cancer Research, 54***,** *4756-4759.*

*RAJENDRA, R., MALEGAONKAR, D., PUNGALIYA, P., MARSHALL, H., RASHEED, Z., BROWNELL, J., LIU, L. F., LUTZKER, S., SALEEM, A. & RUBIN, E. H. 2004. Topors functions as an E3 ubiquitin ligase with specific E2 enzymes and ubiquitinates p53. J Biol Chem, 279***,** *36440-4.*

*RAKYAN, V. K., DOWN, T. A., MASLAU, S., ANDREW, T., YANG, T. P., BEYAN, H., WHITTAKER, P., MCCANN, O. T., FINER, S., VALDES, A. M., LESLIE, R. D., DELOUKAS, P. & SPECTOR, T. D. 2010. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res, 20***,** *434-9.*

*RAKYAN, V. K., DOWN, T. A., THORNE, N. P., FLICEK, P., KULESHA, E., GRAF, S., TOMAZOU, E. M., BACKDAHL, L., JOHNSON, N., HERBERTH, M., HOWE, K. L., JACKSON, D. K., MIRETTI, M. M., FIEGLER, H., MARIONI, J. C., BIRNEY, E., HUBBARD, T. J., CARTER, N. P., TAVARE, S. & BECK, S. 2008. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res, 18***,** *1518-29.*

*RAMSAHOYE, B. H., BINISZKIEWICZ, D., LYKO, F., CLARK, V., BIRD, A. P. & JAENISCH, R. 2000. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences of the United States of America, 97***,** *5237-42.*

*RAPONI, M., DOSSEY, L., JATKOE, T., WU, X., CHEN, G., FAN, H. & BEER, D. G. 2009. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. Cancer Res, 69***,** *5776-83.*

*RAUCH, T., WANG, Z., ZHANG, X., ZHONG, X., WU, X., LAU, S. K., KERNSTINE, K. H., RIGGS, A. D. & PFEIFER, G. P. 2007. Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. Proceedings of the National Academy of Sciences of the United States of America, 104***,** *5527-32.*

RECKOVA, M., ROSENGARTEN, C., DEALMEIDA, A., STANLEY, C. P., WESSELS, A., GOURDIE, R. G., THOMPSON, R. P. & SEDMERA, D. 2003. Hemodynamics Is a Key Epigenetic Factor in Development of the Cardiac Conduction System. Circ Res, 93, 77-85.

RECZKO, M., MARAGKAKIS, M., ALEXIOU, P., GROSSE, I. & HATZIGEORGIOU, A. G. 2012. Functional microRNA targets in protein coding sequences. Bioinformatics, 28, 771-6.

REDFERN, A. D., COLLEY, S. M., BEVERIDGE, D. J., IKEDA, N., EPIS, M. R., LI, X., FOULDS, C. E., STUART, L. M., BARKER, A., RUSSELL, V. J., RAMSAY, K., KOBELKE, S. J., LI, X., HATCHELL, E. C., PAYNE, C., GILES, K. M., MESSINEO, A., GATIGNOL, A., LANZ, R. B., O'MALLEY, B. W. & LEEDMAN, P. J. 2013. RNA-induced silencing complex (RISC) Proteins PACT, TRBP, and Dicer are SRA binding nuclear receptor coregulators. Proc Natl Acad Sci U S A, 110, 6536-41.

REHEMTULLA, A. 2010. Dinosaurs and ancient civilizations: reflections on the treatment of cancer. Neoplasia, 12, 957-68.

REHMAN, I., QUINN, A. G., HEALY, E. & REES, J. L. 1994. High frequency of loss of heterozygosity in actinic keratoses, a usually benign disease. Lancet, 344, 788-9.

REHMAN, I., QUINN, A. G., TAKATA, M., TAYLOR, A. E. & REES, J. L. 1997. Low frequency of allelic loss in skin tumours from immunosuppressed individuals. Br J Cancer, 76, 757-9.

REICHRATH, J. & NURNBERG, B. 2008. Solar UV-radiation, vitamin D and skin cancer surveillance in organ transplant recipients (OTRs). Adv Exp Med Biol, 624, 203-14.

REIK, W., DEAN, W. & WALTER, J. 2001. Epigenetic reprogramming in mammalian development. Science, 293, 1089-93.

REINERS, C., DEMIDCHIK, Y. E., DROZD, V. M. & BIKO, J. 2008. Thyroid cancer in infants and adolescents after Chernobyl. Minerva Endocrinol, 33, 381-95.

REINHART, B. J., SLACK, F. J., BASSON, M., PASQUINELLI, A. E., BETTINGER, J. C., ROUGVIE, A. E., HORVITZ, H. R. & RUVKUN, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature, 403, 901-6.

REYA, T., MORRISON, S. J., CLARKE, M. F. & WEISSMAN, I. L. 2001. Stem cells, cancer, and cancer stem cells. Nature, 414, 105-11.

RICART, A. D. 2011. Immunoconjugates against solid tumors: mind the gap. Clin Pharmacol Ther, 89, 513-23.

RIED, T. 2009. Homage to Theodor Boveri (1862–1915): Boveri's theory of cancer as a disease of the chromosomes, and the landscape of genomic imbalances in human carcinomas. Environmental and Molecular Mutagenesis, 50, 593-601.

RIGGI, N., SUVA, M. L., SUVA, D., CIRONI, L., PROVERO, P., TERCIER, S., JOSEPH, J. M., STEHLE, J. C., BAUMER, K., KINDLER, V. & STAMENKOVIC, I. 2008. EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. Cancer Res, 68, 2176-85.

RIKER, A., ENKEMANN, S., FODSTAD, O., LIU, S., REN, S., MORRIS, C., XI, Y., HOWELL, P., METGE, B., SAMANT, R., SHEVDE, L., LI, W., ESCHRICH, S., DAUD, A., JU, J. & MATTA, J. 2008. The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. BMC Medical Genomics, 1, 13.

ROBINTON, D. A. & DALEY, G. Q. 2012. The promise of induced pluripotent stem cells in research and therapy. Nature, 481, 295-305.

RODRIGUEZ, J., FRIGOLA, J., VENDRELL, E., RISQUES, R. A., FRAGA, M. F., MORALES, C., MORENO, V., ESTELLER, M., CAPELLA, G., RIBAS, M. & PEINADO, M. A. 2006. Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. Cancer Res, 66, 8462-9468.

RODUST, P. M., STOCKFLETH, E., ULRICH, C., LEVERKUS, M. & EBERLE, J. 2009. UV-induced squamous cell carcinoma--a role for antiapoptotic signalling pathways. Br J Dermatol, 161 Suppl 3, 107-15.

ROKAVEC, M., WU, W. & LUO, J. L. 2012. IL6-mediated suppression of miR-200c directs constitutive activation of inflammatory signaling circuit driving transformation and tumorigenesis. Mol Cell, 45, 777-89.

ROLLISON, D. E., GIULIANO, A. R., MESSINA, J. L., FENSKE, N. A., CHERPELIS, B. S., SONDAK, V. K., ROETZHEIM, R. G., IANNACONE, M. R., MICHAEL, K. M., GHEIT, T., WATERBOER, T., TOMMASINO, M. & PAWLITA, M. 2012. Case–control Study of Merkel Cell Polyomavirus Infection and Cutaneous Squamous Cell Carcinoma. Cancer Epidemiology Biomarkers & Prevention, 21, 74-81.

ROOK, A. & BURNS, T. 2010. Rook's textbook of dermatology, Chichester, Wiley-Blackwell.

ROORDA, B., TER ELST, A., KAMPS, W. & DE BONT, E. 2009. Bone marrow-derived cells and tumor growth: contribution of bone marrow-derived cells to tumor micro-environments with special focus on mesenchymal stem cells. Crit Rev Oncol Hematol, 69, 187 - 98.

ROSASCO-NITCHER, S. E., LAN, W., KHORASANIZADEH, S. & STUKENBERG, P. T. 2008. Centromeric Aurora-B activation requires TD-60, microtubules, and substrate priming phosphorylation. Science, 319, 469-72.

ROSEN, J. M. & JORDAN, C. T. 2009. The increasing complexity of the cancer stem cell paradigm. Science, 324, 1670-3.

ROUS, P. 1911. A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. J Exp Med, 13, 397-411.

RUBIN, H. 2011a. The early history of tumor virology: Rous, RIF, and RAV. Proceedings of the National Academy of Sciences of the United States of America, 108, 14389-96.

RUBIN, H. 2011b. Fields and field cancerization: the preneoplastic origins of cancer: asymptomatic hyperplastic fields are precursors of neoplasia, and their progression to tumors can be tracked by saturation density in culture. Bioessays, 33, 224-31.

RUBIO, D., GARCIA, S., PAZ, M. F., DE LA CUEVA, T., LOPEZ-FERNANDEZ, L. A., LLOYD, A. C., GARCIA-CASTRO, J. & BERNAD, A. 2008. Molecular characterization of spontaneous mesenchymal stem cell transformation. PLoS One, 3, e1398.

SAFTIG, P. & REISS, K. 2011. The "A Disintegrin And Metalloproteases" ADAM10 and ADAM17: Novel drug targets with therapeutic potential? Eur J Cell Biol, 90, 527-35.

SALASCHE, S. J. 2000. Epidemiology of actinic keratoses and squamous cell carcinoma. J Am Acad Dermatol, 42, 4-7.

SALDANA-MEYER, R. & RECILLAS-TARGA, F. 2011. Transcriptional and epigenetic regulation of the p53 tumor suppressor gene. Epigenetics, 6, 1068-77.

SALGADO, R., TOLL, A., ALAMEDA, F., BARO, T., MARTIN-EZQUERRA, G., SANMARTIN, O., MARTORELL-CALATAYUD, A., SALIDO, M., ALMENAR, S., SOLE, F., PUJOL, R. M. & ESPINET, B. 2010. CKS1B amplification is a frequent event in cutaneous squamous cell carcinoma with aggressive clinical behaviour. Genes Chromosomes Cancer, 49, 1054-61.

SAMPOGNA, F., BAVINCK, J. N. B., PAWLITA, M., ABENI, D., HARWOOD, C. A., PROBY, C. M., FELTKAMP, M. C. W., EUVRARD, S., NALDI, L., NEALE, R. E., NINDL, I., PFISTER, H., QUINT, W. G. V., WATERBOER, T. & GROUP, T. E.-H.-U.-C. 2012. Factors associated with the seroprevalence of 26 cutaneous and two genital human papillomavirus types in organ transplant patients. Journal of General Virology, 93, *165-174.*

SAND, M., SKRYGAN, M., GEORGAS, D., SAND, D., HAHN, S. A., GAMBICHLER, T., ALTMEYER, P. & BECHARA, F. G. 2012a. Microarray analysis of microRNA expression in cutaneous squamous cell carcinoma. J Dermatol Sci, 68, *119-26.*

SAND, M., SKRYGAN, M., SAND, D., GEORGAS, D., HAHN, S. A., GAMBICHLER, T., ALTMEYER, P. & BECHARA, F. G. 2012b. Expression of microRNAs in basal cell carcinoma. Br J Dermatol, 167, *847-55.*

SANDOVAL, J., HEYN, H., MORAN, S., SERRA-MUSACH, J., PUJANA, M. A., BIBIKOVA, M. & ESTELLER, M. 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics, 6, *692-702.*

SASLOW, D., CASTLE, P. E., COX, J. T., DAVEY, D. D., EINSTEIN, M. H., FERRIS, D. G., GOLDIE, S. J., HARPER, D. M., KINNEY, W., MOSCICKI, A.-B., NOLLER, K. L., WHEELER, C. M., ADES, T., ANDREWS, K. S., DOROSHENK, M. K., KAHN, K. G., SCHMIDT, C., SHAFEY, O., SMITH, R. A., PARTRIDGE, E. E. & GARCIA, F. 2007. American Cancer Society Guideline for Human Papillomavirus (HPV) Vaccine Use to Prevent Cervical Cancer and Its Precursors. CA: A Cancer Journal for Clinicians, 57, *7-28.*

SASTRE-GARAU, X., PETER, M., AVRIL, M.-F., LAUDE, H., COUTURIER, J., ROZENBERG, F., ALMEIDA, A., BOITIER, F., CARLOTTI, A., COUTURAUD, B. & DUPIN, N. 2009. Merkel cell carcinoma of the skin: pathological and molecular evidence for a causative role of MCV in oncogenesis. The Journal of Pathology, 218, *48-56.*

SATHYANARAYANA, U. G., MOORE, A. Y., LI, L., PADAR, A., MAJMUDAR, K., STASTNY, V., MAKARLA, P., SUZUKI, M., MINNA, J. D., FENG, Z. & GAZDAR, A. F. 2007. Sun exposure related methylation in malignant and non-malignant skin lesions. Cancer letters, 245, *112-20.*

SATO, M., NISHIGORI, C., LU, Y., ZGHAL, M., YAGI, T. & TAKEBE, H. 1994. Far less frequent mutations in ras genes than in the p53 gene in skin tumors of xeroderma pigmentosum patients. Molecular Carcinogenesis, 11, *98-105.*

SATO, M., TAKAHASHI, K., NAGAYAMA, K., ARAI, Y., ITO, N., OKADA, M., MINNA, J. D., YOKOTA, J. & KOHNO, T. 2005. Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. Genes Chromosomes Cancer, 44, *405-14.*

SATO, T., KANEDA, A., TSUJI, S., ISAGAWA, T., YAMAMOTO, S., FUJITA, T., YAMANAKA, R., TANAKA, Y., NUKIWA, T., MARQUEZ, V. E., ISHIKAWA, Y., ICHINOSE, M. & ABURATANI, H. 2013. PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. Sci. Rep., 3.

SAUERMANN, M., SAHIN, O., SULTMANN, H., HAHNE, F., BLASZKIEWICZ, S., MAJETY, M., ZATLOUKAL, K., FUZESI, L., POUSTKA, A., WIEMANN, S. & ARLT, D. 2008. Reduced expression of vacuole membrane protein 1 affects the invasion capacity of tumor cells. Oncogene, 27, *1320-6.*

SAULNIER, A., VAISSIERE, T., YUE, J., SIOUDA, M., MALFROY, M., ACCARDI, R., CREVEAUX, M., SEBASTIAN, S., SHAHZAD, N., GHEIT, T., HUSSAIN, I., TORRENTE, M., MAFFINI, F. A., CALABRESE, L., CHIESA, F., CUENIN, C., SHUKLA, R., FATHALLAH, I., MATOS, E., DAUDT, A., KOIFMAN, S., WUNSCH-FILHO, V., MENEZES, A. M., CURADO, M. P., ZARIDZE, D., BOFFETTA, P., BRENNAN, P., TOMMASINO, M., HERCEG, Z. & SYLLA, B. S. 2012.

Inactivation of the putative suppressor gene DOK1 by promoter hypermethylation in primary human cancers. Int J Cancer, 130, *2484-94.*

SAVAGE, D. C. 1977. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol, 31, *107-33.*

SCHACHT, V., DADRAS, S. S., JOHNSON, L. A., JACKSON, D. G., HONG, Y. K. & DETMAR, M. 2005. Up-regulation of the lymphatic marker podoplanin, a mucin-type transmembrane glycoprotein, in human squamous cell carcinomas and germ cell tumors. Am J Pathol, 166, *913-21.*

SCHIPMANN, S., WERMKER, K., SCHULZE, H. J., KLEINHEINZ, J. & BRUNNER, G. 2014. Cutaneous and oral squamous cell carcinoma-dual immunosuppression via recruitment of FOXP3 regulatory T cells and endogenous tumour FOXP3 expression? J Craniomaxillofac Surg.

SCHMITT, J. V. & MIOT, H. A. 2012. Actinic keratosis: a clinical and epidemiological revision. An Bras Dermatol, 87, *425-34.*

SCHUETTENGRUBER, B., CHOURROUT, D., VERVOORT, M., LEBLANC, B. & CAVALLI, G. 2007. Genome Regulation by Polycomb and Trithorax Proteins. Cell, 128, *735-745.*

SCHWARTZ, J. L., BICHAKJIAN, C. K., LOWE, L., GRIFFITH, K. A., FROHM, M. L., FULLEN, D. R., HAYMAN, J. A., LAO, C. D., SHAH, K. S., MCLEAN, S. A., BRADFORD, C. R., JOHNSON, T. M. & WONG, S. L. 2013. Clinicopathologic Features of Primary Merkel Cell Carcinoma: A Detailed Descriptive Analysis of a Large Contemporary Cohort. Dermatologic Surgery, 39, *1009-1016.*

SCHWARTZ, S. D., HUBSCHMAN, J. P., HEILWELL, G., FRANCO-CARDENAS, V., PAN, C. K., OSTRICK, R. M., MICKUNAS, E., GAY, R., KLIMANSKAYA, I. & LANZA, R. 2012. Embryonic stem cell trials for macular degeneration: a preliminary report. Lancet, 379, *713-20.*

SCOLA, N., WIELAND, U., SILLING, S., ALTMEYER, P., STÜCKER, M. & KREUTER, A. 2012. Prevalence of human polyomaviruses in common and rare types of non-Merkel cell carcinoma skin cancer. British Journal of Dermatology, 167, *1315-1320.*

SEN, G. L., BOXER, L. D., WEBSTER, D. E., BUSSAT, R. T., QU, K., ZARNEGAR, B. J., JOHNSTON, D., SIPRASHVILI, Z. & KHAVARI, P. A. 2012. ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. Dev Cell, 22, *669-77.*

SEN, G. L., WEBSTER, D. E., BARRAGAN, D. I., CHANG, H. Y. & KHAVARI, P. A. 2008. Control of differentiation in a self-renewing mammalian tissue by the histone demethylase JMJD3. Genes Dev, 22, *1865-70.*

SEO, G. J., CHEN, C. J. & SULLIVAN, C. S. 2009. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression. Virology, 383, *183-7.*

SEVERSON, P. L., TOKAR, E. J., VRBA, L., WAALKES, M. P. & FUTSCHER, B. W. 2013. Coordinate H3K9 and DNA methylation silencing of ZNFs in toxicant-induced malignant transformation. Epigenetics, 8, *1080-1088.*

SHANN, Y. J., CHENG, C., CHIAO, C. H., CHEN, D. T., LI, P. H. & HSU, M. T. 2008. Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines. Genome Res, 18, *791-801.*

SHARP, C. P., NORJA, P., ANTHONY, I., BELL, J. E. & SIMMONDS, P. 2009. Reactivation and mutation of newly discovered WU, KI, and Merkel cell carcinoma polyomaviruses in immunosuppressed individuals. The Journal of infectious diseases, 199, *398-404.*

*SHE, Q.-B., CHEN, N., BODE, A. M., FLAVELL, R. A. & DONG, Z. 2002. Deficiency of c-Jun-NH2-terminal Kinase-1 in Mice Enhances Skin Tumor Development by 12-O-Tetradecanoylphorbol-13-Acetate. Cancer Research, 62***,** *1343-1348.*

*SHEN, Q., JIN, H. & WANG, X. 2013. Epidermal stem cells and their epigenetic regulation. Int J Mol Sci, 14***,** *17861-80.*

*SHERR, C. J. 1996. Cancer cell cycles. Science, 274***,** *1672-7.*

*SHIH, C. C., FORMAN, S. J., CHU, P. & SLOVAK, M. 2007. Human embryonic stem cells are prone to generate primitive, undifferentiated tumors in engrafted human fetal tissues in severe combined immunodeficient mice. Stem Cells Dev, 16***,** *893-902.*

*SHIIBA, M., NOMURA, H., SHINOZUKA, K., SAITO, K., KOUZU, Y., KASAMATSU, A., SAKAMOTO, Y., MURANO, A., ONO, K., OGAWARA, K., UZAWA, K. & TANZAWA, H. 2010. Down-regulated expression of SERPIN genes located on chromosome 18q21 in oral squamous cell carcinomas. Oncol Rep, 24***,** *241-9.*

*SHIMOMURA, Y., AOKI, N., ROGERS, M. A., LANGBEIN, L., SCHWEIZER, J. & ITO, M. 2002. hKAP1.6 and hKAP1.7, two novel human high sulfur keratin-associated proteins are expressed in the hair follicle cortex. J Invest Dermatol, 118***,** *226-31.*

*SHORT, K., WIRADJAJA, F. & SMYTH, I. 2007. Let's stick together: the role of the Fras1 and Frem proteins in epidermal adhesion. IUBMB Life, 59***,** *427-35.*

*SHUDA, M., FENG, H., KWUN, H. J., ROSEN, S. T., GJOERUP, O., MOORE, P. S. & CHANG, Y. 2008. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proceedings of the National Academy of Sciences of the United States of America, 105***,** *16272-7.*

*SHYU, A. B., WILKINSON, M. F. & VAN HOOF, A. 2008. Messenger RNA regulation: to translate or to degrade. EMBO J, 27***,** *471-81.*

*SIEGMUND, K. D., MARJORAM, P., WOO, Y. J., TAVARE, S. & SHIBATA, D. 2009. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America, 106***,** *4828-33.*

*SIHTO, H., KUKKO, H., KOLJONEN, V., SANKILA, R., BOHLING, T. & JOENSUU, H. 2011. Merkel cell polyomavirus infection, large T antigen, retinoblastoma protein and outcome in Merkel cell carcinoma. Clin Cancer Res, 17***,** *4806-13.*

*SINGH, N., MORLOCK, H. & HANES, S. D. 2011. The Bin3 RNA methyltransferase is required for repression of caudal translation in the Drosophila embryo. Dev Biol, 352***,** *104-15.*

*SNIPPERT, H. J., HAEGEBARTH, A., KASPER, M., JAKS, V., VAN ES, J. H., BARKER, N., VAN DE WETERING, M., VAN DEN BORN, M., BEGTHEL, H., VRIES, R. G., STANGE, D. E., TOFTGARD, R. & CLEVERS, H. 2010. Lgr6 marks stem cells in the hair follicle that generate all cell lineages of the skin. Science, 327***,** *1385-9.*

*SONG, Y. X., YUE, Z. Y., WANG, Z. N., XU, Y. Y., LUO, Y., XU, H. M., ZHANG, X., JIANG, L., XING, C. Z. & ZHANG, Y. 2011. MicroRNA-148b is frequently down-regulated in gastric cancer and acts as a tumor suppressor by inhibiting cell proliferation. Mol Cancer, 10***,** *1.*

*SONKOLY, E., LOVEN, J., XU, N., MEISGEN, F., WEI, T., BRODIN, P., JAKS, V., KASPER, M., SHIMOKAWA, T., HARADA, M., HEILBORN, J., HEDBLAD, M. A., HIPPE, A., GRANDER, D., HOMEY, B., ZAPHIROPOULOS, P. G., ARSENIAN-HENRIKSSON, M., STAHLE, M. & PIVARCSI, A. 2012. MicroRNA-203 functions as a tumor suppressor in basal cell carcinoma. Oncogenesis, 1***,** *e3.*

*SONKOLY, E., WEI, T., JANSON, P. C., SAAF, A., LUNDEBERG, L., TENGVALL-LINDER, M., NORSTEDT, G., ALENIUS, H., HOMEY, B., SCHEYNIUS, A., STAHLE, M. & PIVARCSI, A. 2007. MicroRNAs: novel regulators involved in the pathogenesis of psoriasis? PLoS One, 2***,** *e610.*

*SOUFIR, N., MOLES, J. P., VILMER, C., MOCH, C., VEROLA, O., RIVET, J., TESNIERE, A., DUBERTRET, L. & BASSET-SEGUIN, N. 1999. P16 UV mutations in human skin epithelial tumors. Oncogene, 18***,** *5477-81.*

*SPERGER, J. M., CHEN, X., DRAPER, J. S., ANTOSIEWICZ, J. E., CHON, C. H., JONES, S. B., BROOKS, J. D., ANDREWS, P. W., BROWN, P. O. & THOMSON, J. A. 2003. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. Proceedings of the National Academy of Sciences of the United States of America, 100***,** *13350-5.*

*SPROUL, D., KITCHEN, R. R., NESTOR, C. E., DIXON, J. M., SIMS, A. H., HARRISON, D. J., RAMSAHOYE, B. H. & MEEHAN, R. R. 2012. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome Biol, 13***,** *R84.*

*SPURGEON, M. E. & LAMBERT, P. F. 2013. Merkel cell polyomavirus: a newly discovered human virus with oncogenic potential. Virology, 435***,** *118-30.*

*STAGG, J. 2008. Mesenchymal stem cells in cancer. Stem Cell Rev, 4***,** *119-24.*

*STALLINGS, R. L., NAIR, P., MARIS, J. M., CATCHPOOLE, D., MCDERMOTT, M., O'MEARA, A. & BREATNACH, F. 2006. High-resolution analysis of chromosomal breakpoints and genomic instability identifies PTPRD as a candidate tumor suppressor gene in neuroblastoma. Cancer Res, 66***,** *3673-80.*

*STATHIS, A., CHINI, C., BERTONI, F., PROSERPIO, I., CAPELLA, C., MAZZUCCHELLI, L., PEDRINIS, E., CAVALLI, F., PINOTTI, G. & ZUCCA, E. 2009. Long-term outcome following Helicobacter pylori eradication in a retrospective study of 105 patients with localized gastric marginal zone B-cell lymphoma of MALT type. Annals of Oncology, 20***,** *1086-1093.*

STATISTICS, O. F. N. 2010. Cancer: One in four people die from cancer [Online]. Available: http://www.statistics.gov.uk/cci/nugget.asp?id=915 *[Accessed 19.5.2011 2011].*

*STEBEN, M. & DUARTE-FRANCO, E. 2007. Human papillomavirus infection: epidemiology and pathophysiology. Gynecol Oncol, 107***,** *S2-5.*

*STEINMANN, K., SANDNER, A., SCHAGDARSURENGIN, U. & DAMMANN, R. H. 2009. Frequent promoter hypermethylation of tumor-related genes in head and neck squamous cell carcinoma. Oncology reports, 22***,** *1519-26.*

*STERN, R. S. 2012. The risk of squamous cell and basal cell cancer associated with psoralen and ultraviolet A therapy: A 30-year prospective study. Journal of the American Academy of Dermatology, 66***,** *553-562.*

*STOFF, B., SALISBURY, C., PARKER, D. & O'REILLY ZWALD, F. 2010. Dermatopathology of skin cancer in solid organ transplant recipients. Transplantation reviews, 24***,** *172-89.*

STOREY, A. & SIMMONDS, M. 2009. Interaction Between Ultraviolet Radiation and Human Papillomavirus. In: STOCKFLETH, E. & ULRICH, C. (eds.) Skin Cancer after Organ Transplantation. Springer US.

*STRATTON, M. R. 2011. Exploring the genomes of cancer cells: progress and promise. Science, 331***,** *1553-8.*

STRAUSSMAN, R., MORIKAWA, T., SHEE, K., BARZILY-ROKNI, M., QIAN, Z. R., DU, J., DAVIS, A., MONGARE, M. M., GOULD, J., FREDERICK, D. T., COOPER, Z. A., CHAPMAN, P. B., SOLIT, D. B., RIBAS, A., LO, R. S., FLAHERTY, K. T., OGINO, S., WARGO, J. A. & GOLUB, T. R. 2012.

*Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. Nature, 487***,** *500-4.*

*STRAZISAR, M., ROTT, T. & GLAVAC, D. 2009. Frequent polymorphic variations but rare tumour specific mutations of the S100A2 on 1q21 in non-small cell lung cancer. Lung Cancer, 63***,** *354-9.*

*STROUHAL, E. 1978. Ancient Egyptian case of carcinoma. Bull N Y Acad Med, 54***,** *290-302.*

*SU, F., VIROS, A., MILAGRE, C., TRUNZER, K., BOLLAG, G., SPLEISS, O., REIS-FILHO, J. S., KONG, X., KOYA, R. C., FLAHERTY, K. T., CHAPMAN, P. B., KIM, M. J., HAYWARD, R., MARTIN, M., YANG, H., WANG, Q., HILTON, H., HANG, J. S., NOE, J., LAMBROS, M., GEYER, F., DHOMEN, N., NICULESCU-DUVAZ, I., ZAMBON, A., NICULESCU-DUVAZ, D., PREECE, N., ROBERT, L., OTTE, N. J., MOK, S., KEE, D., MA, Y., ZHANG, C., HABETS, G., BURTON, E. A., WONG, B., NGUYEN, H., KOCKX, M., ANDRIES, L., LESTINI, B., NOLOP, K. B., LEE, R. J., JOE, A. K., TROY, J. L., GONZALEZ, R., HUTSON, T. E., PUZANOV, I., CHMIELOWSKI, B., SPRINGER, C. J., MCARTHUR, G. A., SOSMAN, J. A., LO, R. S., RIBAS, A. & MARAIS, R. 2012a. RAS Mutations in Cutaneous Squamous-Cell Carcinomas in Patients Treated with BRAF Inhibitors. New England Journal of Medicine, 366***,** *207-215.*

*SU, F., VIROS, A., MILAGRE, C., TRUNZER, K., BOLLAG, G., SPLEISS, O., REIS-FILHO, J. S., KONG, X., KOYA, R. C., FLAHERTY, K. T., CHAPMAN, P. B., KIM, M. J., HAYWARD, R., MARTIN, M., YANG, H., WANG, Q., HILTON, H., HANG, J. S., NOE, J., LAMBROS, M., GEYER, F., DHOMEN, N., NICULESCU-DUVAZ, I., ZAMBON, A., NICULESCU-DUVAZ, D., PREECE, N., ROBERT, L., OTTE, N. J., MOK, S., KEE, D., MA, Y., ZHANG, C., HABETS, G., BURTON, E. A., WONG, B., NGUYEN, H., KOCKX, M., ANDRIES, L., LESTINI, B., NOLOP, K. B., LEE, R. J., JOE, A. K., TROY, J. L., GONZALEZ, R., HUTSON, T. E., PUZANOV, I., CHMIELOWSKI, B., SPRINGER, C. J., MCARTHUR, G. A., SOSMAN, J. A., LO, R. S., RIBAS, A. & MARAIS, R. 2012b. RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. N Engl J Med, 366***,** *207-15.*

*SUNDARAM, M. V. 2006. RTK/Ras/MAPK signaling. WormBook***,** *1-19.*

*SUOMELA, S., KOLJONEN, V., SKOOG, T., KUKKO, H., BOHLING, T. & SAARIALHO-KERE, U. 2009. Expression of MMP-10, MMP-21, MMP-26, and MMP-28 in Merkel cell carcinoma. Virchows Arch, 455***,** *495-503.*

*SVENSSON MANSSON, S., REIS-FILHO, J. & LANDBERG, G. 2007. Transcriptional upregulation and unmethylation of the promoter region of p16 in invasive basal cell carcinoma cells and partial co-localization with the gamma 2 chain of laminin-332. The Journal of pathology, 212***,** *102-11.*

*TABS, S. & AVCI, O. 2004. Induction of the differentiation and apoptosis of tumor cells in vivo with efficiency and selectivity. Eur J Dermatol, 14***,** *96-102.*

*TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science, 324***,** *930-935.*

*TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 131***,** *861-72.*

*TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 126***,** *663-76.*

TAKAHASHI, Y., SAWADA, G., KURASHIGE, J., MATSUMURA, T., UCHI, R., UEO, H., ISHIBASHI, M., TAKANO, Y., AKIYOSHI, S., IWAYA, T., EGUCHI, H., SUDO, T., SUGIMACHI, K., YAMAMOTO, H., DOKI, Y., MORI, M. & MIMORI, K. 2013. Tumor-derived tenascin-C promotes the epithelial-mesenchymal transition in colorectal cancer cells. Anticancer Res, 33, *1927-34.*

TAKEUCHI, T., LIANG, S. B., MATSUYOSHI, N., ZHOU, S., MIYACHI, Y., SONOBE, H. & OHTSUKI, Y. 2002a. Loss of T-cadherin (CDH13, H-cadherin) expression in cutaneous squamous cell carcinoma. Lab Invest, 82, *1023-9.*

TAKEUCHI, T., LIANG, S. B. & OHTSUKI, Y. 2002b. Downregulation of expression of a novel cadherin molecule, T-cadherin, in basal cell carcinoma of the skin. Molecular carcinogenesis, 35, *173-9.*

TANEMURA, A., NAGASAWA, T., INUI, S. & ITAMI, S. 2005. LRIG-1 provides a novel prognostic predictor in squamous cell carcinoma of the skin: immunohistochemical analysis for 38 cases. Dermatol Surg, 31, *423-30.*

TANG, J. Y., MACKAY-WIGGAN, J. M., ASZTERBAUM, M., YAUCH, R. L., LINDGREN, J., CHANG, K., COPPOLA, C., CHANANA, A. M., MARJI, J., BICKERS, D. R. & EPSTEIN, E. H., JR. 2012. Inhibiting the hedgehog pathway in patients with the basal-cell nevus syndrome. N Engl J Med, 366, *2180-8.*

TANG, P., HUANG, H., CHANG, J., ZHAO, G. F., LU, M. L. & WANG, Y. 2013. Increased expression of DLX2 correlates with advanced stage of gastric adenocarcinoma. World J Gastroenterol, 19, *2697-703.*

TANNOCK, I. 2005. The basic science of oncology, New York London, McGraw-Hill,Medical Publishing Division.

TATENO, H., MATSUSHIMA, A., HIEMORI, K., ONUMA, Y., ITO, Y., HASEHIRA, K., NISHIMURA, K., OHTAKA, M., TAKAYASU, S., NAKANISHI, M., IKEHARA, Y., NAKANISHI, M., OHNUMA, K., CHAN, T., TOYODA, M., AKUTSU, H., UMEZAWA, A., ASASHIMA, M. & HIRABAYASHI, J. 2013. Podocalyxin is a glycoprotein ligand of the human pluripotent stem cell-specific probe rBC2LCN. Stem Cells Transl Med, 2, *265-73.*

TATZELT, J. & SCHATZL, H. M. 2007. Molecular basis of cerebral neurodegeneration in prion diseases. FEBS J, 274, *606-11.*

TAYLOR, K. H., PENA-HERNANDEZ, K. E., DAVIS, J. W., ARTHUR, G. L., DUFF, D. J., SHI, H., RAHMATPANAH, F. B., SJAHPUTERA, O. & CALDWELL, C. W. 2007. Large-scale CpG methylation analysis identifies novel candidate genes and reveals methylation hotspots in acute lymphoblastic leukemia. Cancer Res, 67, *2617-25.*

TEMAN, C. J., TRIPP, S. R., PERKINS, S. L. & DUNCAVAGE, E. J. 2011. Merkel cell polyomavirus (MCPyV) in chronic lymphocytic leukemia/small lymphocytic lymphoma. Leuk Res, 35, *689-92.*

THOMPSON, L. L., GUPPY, B. J., SAWCHUK, L., DAVIE, J. R. & MCMANUS, K. J. 2013. Regulation of chromatin structure via histone post-translational modification and the link to carcinogenesis. Cancer Metastasis Rev, 32, *363-76.*

THOMPSON, S. C., JOLLEY, D. & MARKS, R. 1993. Reduction of Solar Keratoses by Regular Sunscreen Use. New England Journal of Medicine, 329, *1147-1151.*

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. Science, 282, *1145-7.*

*THOMSON, J. P., HUNTER, J. M., LEMPIÄINEN, H., MÜLLER, A., TERRANOVA, R., MOGGS, J. G. & MEEHAN, R. R. 2013. Dynamic changes in 5-hydroxymethylation signatures underpin early and late events in drug exposed liver. Nucleic Acids Research, 41***,** *5639-5654.*

*TIERNEY, P., FERGUSON, J., IBBOTSON, S., DAWE, R., EADIE, E. & MOSELEY, H. 2013. Nine out of 10 sunbeds in England emit ultraviolet radiation levels that exceed current safety limits. British Journal of Dermatology, 168***,** *602-608.*

*TISCHKOWITZ, M., BRUNET, J. S., BEGIN, L. R., HUNTSMAN, D. G., CHEANG, M. C., AKSLEN, L. A., NIELSEN, T. O. & FOULKES, W. D. 2007. Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. BMC Cancer, 7***,** *134.*

*TIWARI, N., MEYER-SCHALLER, N., ARNOLD, P., ANTONIADIS, H., PACHKOV, M., VAN NIMWEGEN, E. & CHRISTOFORI, G. 2013. Klf4 is a transcriptional regulator of genes critical for EMT, including Jnk1 (Mapk8). PLoS One, 8***,** *e57329.*

*TOKER, C. 1972. Trabecular carcinoma of the skin. Archives of dermatology, 105***,** *107-10.*

*TOLL, A., SALGADO, R., YÉBENES, M., MARTÍN-EZQUERRA, G., GILABERTE, M., BARÓ, T., SOLÉ, F., ALAMEDA, F., ESPINET, B. & PUJOL, R. M. 2009. MYC gene numerical aberrations in actinic keratosis and cutaneous squamous cell carcinoma. British Journal of Dermatology, 161***,** *1112-1118.*

*TOLSTOV, Y. L., KNAUER, A., CHEN, J. G., KENSLER, T. W., KINGSLEY, L. A., MOORE, P. S. & CHANG, Y. 2011. Asymptomatic primary Merkel cell polyomavirus infection among adults. Emerging infectious diseases, 17***,** *1371-80.*

*TOMIZAWA, S.-I., KOBAYASHI, H., WATANABE, T., ANDREWS, S., HATA, K., KELSEY, G. & SASAKI, H. 2011. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development, 138***,** *811-820.*

*TOMMASI, S., KARM, D. L., WU, X., YEN, Y. & PFEIFER, G. P. 2009. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. Breast cancer research : BCR, 11***,** *R14.*

*TOYOSHIMA, T., VAIRAKTARIS, E., NKENKE, E., SCHLEGEL, K. A., NEUKAM, F. W. & RIES, J. 2008. Cytokeratin 17 mRNA expression has potential for diagnostic marker of oral squamous cell carcinoma. J Cancer Res Clin Oncol, 134***,** *515-21.*

*TRAKATELLI, M., ULRICH, C., DEL MARMOL, V., EUVRARD, S., STOCKFLETH, E. & ABENI, D. 2007. Epidemiology of nonmelanoma skin cancer (NMSC) in Europe: accurate and comparable data are needed for effective public health monitoring and interventions. Br J Dermatol, 156 Suppl 3***,** *1-7.*

*TRUMBO, C. W., MCCOMAS, K. A. & KANNAOVAKUN, P. 2007. Cancer anxiety and the perception of risk in alarmed communities. Risk Anal, 27***,** *337-50.*

*TSAI, M. S., HWANG, S. M., CHEN, K. D., LEE, Y. S., HSU, L. W., CHANG, Y. J., WANG, C. N., PENG, H. H., CHANG, Y. L., CHAO, A. S., CHANG, S. D., LEE, K. D., WANG, T. H., WANG, H. S. & SOONG, Y. K. 2007. Functional network analysis of the transcriptomes of mesenchymal stem cells derived from amniotic fluid, amniotic membrane, cord blood, and bone marrow. Stem Cells, 25***,** *2511-23.*

*TSUBOTA, Y., OGAWA, T., OYANAGI, J., NAGASHIMA, Y. & MIYAZAKI, K. 2010. Expression of laminin gamma2 chain monomer enhances invasive growth of human carcinoma cells in vivo. Int J Cancer, 127***,** *2031-41.*

TU, W. H., ZHU, C., CLARK, C., CHRISTENSEN, J. G. & SUN, Z. 2010. Efficacy of c-Met inhibitor for advanced prostate cancer. BMC Cancer, 10, *556.*

TYLER, L. N., AI, L., ZUO, C., FAN, C. Y. & SMOLLER, B. R. 2003. Analysis of promoter hypermethylation of death-associated protein kinase and p16 tumor suppressor genes in actinic keratoses and squamous cell carcinomas of the skin. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc, 16, *660-4.*

UCCELLI, A., MORETTA, L. & PISTOIA, V. 2008. Mesenchymal stem cells in health and disease. Nat Rev Immunol, 8, *726 - 36.*

UEDA, R., YOSHIDA, K., KAWASE, T., KAWAKAMI, Y. & TODA, M. 2007. Preferential expression and frequent IgG responses of a tumor antigen, SOX5, in glioma patients. Int J Cancer, 120, *1704-11.*

ULLOA-MONTOYA, F., KIDDER, B. L., PAUWELYN, K. A., CHASE, L. G., LUTTUN, A., CRABBE, A., GERAERTS, M., SHAROV, A. A., PIAO, Y., KO, M. S., HU, W. S. & VERFAILLIE, C. M. 2007. Comparative transcriptome analysis of embryonic and adult stem cells with extended and limited differentiation capacity. Genome Biol, 8, *R163.*

UMBACH, J. L., WANG, K., TANG, S., KRAUSE, P. R., MONT, E. K., COHEN, J. I. & CULLEN, B. R. 2010. Identification of viral microRNAs expressed in human sacral ganglia latently infected with herpes simplex virus 2. J Virol, 84, *1189-92.*

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3--new capabilities and interfaces. Nucleic Acids Res, 40, *e115.*

VALLEJO-TORRES, L., MORRIS, S., KINGE, J. M., POIRIER, V. & VERNE, J. 2013. Measuring current and future cost of skin cancer in England. J Public Health (Oxf).

VAN DER FITS, L., VAN KESTER, M. S., QIN, Y., OUT-LUITING, J. J., SMIT, F., ZOUTMAN, W. H., WILLEMZE, R., TENSEN, C. P. & VERMEER, M. H. 2011. MicroRNA-21 expression in CD4+ T cells is regulated by STAT3 and is pathologically involved in Sezary syndrome. J Invest Dermatol, 131, *762-8.*

VAN DER VELDEN, J. L., GUALA, A. S., LEGGETT, S. E., SLUIMER, J., BADURA, E. C. & JANSSEN-HEININGER, Y. M. 2012. Induction of a mesenchymal expression program in lung epithelial cells by wingless protein (Wnt)/beta-catenin requires the presence of c-Jun N-terminal kinase-1 (JNK1). Am J Respir Cell Mol Biol, 47, *306-14.*

VAN HAREN, R., FELDMAN, D. & SINHA, A. A. 2009. Systematic comparison of nonmelanoma skin cancer microarray datasets reveals lack of consensus genes. British Journal of Dermatology, 161, *1278-1287.*

VAN KESTER, M. S., BALLABIO, E., BENNER, M. F., CHEN, X. H., SAUNDERS, N. J., VAN DER FITS, L., VAN DOORN, R., VERMEER, M. H., WILLEMZE, R., TENSEN, C. P. & LAWRIE, C. H. 2011. miRNA expression profiling of mycosis fungoides. Molecular oncology.

VAN WYNSBERGHE, P. M., CHAN, S. P., SLACK, F. J. & PASQUINELLI, A. E. 2011. Analysis of microRNA expression and function. Methods Cell Biol, 106, *219-52.*

VARLEY, K. E., MUTCH, D. G., EDMONSTON, T. B., GOODFELLOW, P. J. & MITRA, R. D. 2009. Intra-tumor heterogeneity of MLH1 promoter methylation revealed by deep single molecule bisulfite sequencing. Nucleic Acids Res, 37, *4603-12.*

VENANZONI, M. C., ROBINSON, L. R., HODGE, D. R., KOLA, I. & SETH, A. 1996. ETS1 and ETS2 in p53 regulation: spatial separation of ETS binding sites (EBS) modulate protein: DNA interaction. Oncogene, 12, *1199-1204.*

VENTURA, A. & JACKS, T. 2009. MicroRNAs and cancer: short RNAs go a long way. Cell, 136, 586-91.

VENZA, I., VISALLI, M., TRIPODO, B., DE GRAZIA, G., LODDO, S., TETI, D. & VENZA, M. 2010. FOXE1 is a target for aberrant methylation in cutaneous squamous cell carcinoma. The British journal of dermatology, 162, 1093-7.

VILLAR, V., KOCIC, J., BUGARSKI, D., JOVCIC, G. & SANTIBANEZ, J. F. 2010. SKIP is required for TGF-beta1-induced epithelial mesenchymal transition and migration in transformed keratinocytes. FEBS Lett, 584, 4586-92.

VISCIDI, R. P., ROLLISON, D. E., SONDAK, V. K., SILVER, B., MESSINA, J. L., GIULIANO, A. R., FULP, W., AJIDAHUN, A. & RIVANERA, D. 2011. Age-Specific Seroprevalence of Merkel Cell Polyomavirus, BK Virus, and JC Virus. Clinical and Vaccine Immunology, 18, 1737-1743.

VISONE, R., PALLANTE, P., VECCHIONE, A., CIROMBELLA, R., FERRACIN, M., FERRARO, A., VOLINIA, S., COLUZZI, S., LEONE, V., BORBONE, E., LIU, C. G., PETROCCA, F., TRONCONE, G., CALIN, G. A., SCARPA, A., COLATO, C., TALLINI, G., SANTORO, M., CROCE, C. M. & FUSCO, A. 2007. Specific microRNAs are downregulated in human thyroid anaplastic carcinomas. Oncogene, 26, 7590-5.

VISVADER, J. E. 2011. Cells of origin in cancer. Nature, 469, 314-322.

VOGEL, G. & HOLDEN, C. 2008. Stem cells. Ethics questions add to concerns about NIH lines. Science, 321, 756-7.

VOGELSTEIN, B. & KINZLER, K. W. 2004. Cancer genes and the pathways they control. Nat Med, 10, 789-99.

VOGELSTEIN, B., LANE, D. & LEVINE, A. J. 2000. Surfing the p53 network. Nature, 408, 307-310.

VON HOFF, D. D., LORUSSO, P. M., RUDIN, C. M., REDDY, J. C., YAUCH, R. L., TIBES, R., WEISS, G. J., BORAD, M. J., HANN, C. L., BRAHMER, J. R., MACKEY, H. M., LUM, B. L., DARBONNE, W. C., MARSTERS, J. C., JR., DE SAUVAGE, F. J. & LOW, J. A. 2009. Inhibition of the hedgehog pathway in advanced basal-cell carcinoma. N Engl J Med, 361, 1164-72.

WAGNER, J. R., BUSCHE, S., GE, B., KWAN, T., PASTINEN, T. & BLANCHETTE, M. 2014. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol, 15, R37.

WALKER, B. A., LEONE, P. E., JENNER, M. W., LI, C., GONZALEZ, D., JOHNSON, D. C., ROSS, F. M., DAVIES, F. E. & MORGAN, G. J. 2006. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. Blood, 108, 1733-1743.

WALTARI, M., SIHTO, H., KUKKO, H., KOLJONEN, V., SANKILA, R., BOHLING, T. & JOENSUU, H. 2010. Association of Merkel cell polyomavirus infection with tumor p53, KIT, stem cell factor, PDGFR-alpha and survival in Merkel cell carcinoma. International journal of cancer. Journal international du cancer.

WALTER, K., HOLCOMB, T., JANUARIO, T., DU, P., EVANGELISTA, M., KARTHA, N., INIGUEZ, L., SORIANO, R., HUW, L., STERN, H., MODRUSAN, Z., SESHAGIRI, S., HAMPTON, G. M., AMLER, L. C., BOURGON, R., YAUCH, R. L. & SHAMES, D. S. 2012. DNA methylation profiling defines clinically relevant biological subsets of non-small cell lung cancer. Clin Cancer Res, 18, 2360-73.

WANG, J., RAO, S., CHU, J., SHEN, X., LEVASSEUR, D. N., THEUNISSEN, T. W. & ORKIN, S. H. 2006. A protein interaction network for pluripotency of embryonic stem cells. Nature, 444, 364-8.

*WANG, N. J., SANBORN, Z., ARNETT, K. L., BAYSTON, L. J., LIAO, W., PROBY, C. M., LEIGH, I. M., COLLISSON, E. A., GORDON, P. B., JAKKULA, L., PENNYPACKER, S., ZOU, Y., SHARMA, M., NORTH, J. P., VEMULA, S. S., MAURO, T. M., NEUHAUS, I. M., LEBOIT, P. E., HUR, J. S., PARK, K., HUH, N., KWOK, P. Y., ARRON, S. T., MASSION, P. P., BALE, A. E., HAUSSLER, D., CLEAVER, J. E., GRAY, J. W., SPELLMAN, P. T., SOUTH, A. P., ASTER, J. C., BLACKLOW, S. C. & CHO, R. J. 2011. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. Proceedings of the National Academy of Sciences of the United States of America, 108***,** *17761-6.*

WANG, Q., WEN, Y. G., LI, D. P., XIA, J., ZHOU, C. Z., YAN, D. W., TANG, H. M. & PENG, Z. H. 2010. Upregulated INHBA expression is associated with poor survival in gastric cancer. Med Oncol.

*WANG, W., WYCKOFF, J. B., GOSWAMI, S., WANG, Y., SIDANI, M., SEGALL, J. E. & CONDEELIS, J. S. 2007. Coordinated regulation of pathways for enhanced cell motility and chemotaxis is conserved in rat and mouse mammary tumors. Cancer Res, 67***,** *3505-11.*

*WANG, X. & EL NAQA, I. M. 2008. Prediction of both conserved and nonconserved microRNA targets in animals. Bioinformatics, 24***,** *325-32.*

*WANG, X., GHOSH, S. & GUO, S. W. 2001. Quantitative quality control in microarray image processing and data acquisition. Nucleic Acids Res, 29***,** *E75-5.*

*WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet, 40***,** *897-903.*

*WATANABE, J., KANEKO, Y., KUROSUMI, M., KOBAYASHI, Y., SAKAMOTO, M., YOSHIDA, M. A., AKIYAMA, M. & MATSUSHIMA, Y. 2014. High-incidence spontaneous tumors in JF1/Ms mice: relevance of hypomorphic germline mutation and subsequent promoter methylation of Ednrb. J Cancer Res Clin Oncol, 140***,** *99-107.*

*WATSON, R. A. & YEUNG, T. M. 2011. What is the potential of oligodendrocyte progenitor cells to successfully treat human spinal cord injury? BMC Neurol, 11***,** *113.*

*WEBB, A., LI, A. & KAUR, P. 2004. Location and phenotype of human adult keratinocyte stem cells of the skin. Differentiation, 72***,** *387-95.*

WEBER, G. F. 2007. Molecular Mechanisms of Cancer. Dordrecht: Springer.

*WEBER, G. F. & ASHKAR, S. 2000. Molecular mechanisms of tumor dissemination in primary and metastatic brain cancers. Brain Research Bulletin, 53***,** *421-424.*

*WERLING, A. M., DOERFLINGER, Y., BRANDNER, J. M., FUCHS, F., BECKER, J. C., SCHRAMA, D., KURZEN, H., GOERDT, S. & PEITSCH, W. K. 2011. Homo- and heterotypic cell-cell contacts in Merkel cells and Merkel cell carcinomas: heterogeneity and indications for cadherin switching. Histopathology, 58***,** *286-303.*

*WESSELSCHMIDT, R. L. 2011. The teratoma assay: an in vivo assessment of pluripotency. Methods Mol Biol, 767***,** *231-41.*

*WETZELS, C. T., HOEFNAGEL, J. G., BAKKERS, J. M., DIJKMAN, H. B., BLOKX, W. A. & MELCHERS, W. J. 2009. Ultrastructural proof of polyomavirus in Merkel cell carcinoma tumour cells and its absence in small cell carcinoma of the lung. PloS one, 4***,** *e4958.*

*WIELAND, U., SILLING, S., SCOLA, N., POTTHOFF, A., GAMBICHLER, T., BROCKMEYER, N. H., PFISTER, H. & KREUTER, A. 2011. Merkel cell polyomavirus infection in HIV-positive men. Arch Dermatol, 147***,** *401-6.*

WILLIAMS, A. R. & HARE, J. M. 2011. Mesenchymal Stem Cells: Biology, Pathophysiology, Translational Findings, and Therapeutic Implications for Cardiac Disease. Circulation Research, 109, 923-940.

WILLIAMS, K., CHRISTENSEN, J., PEDERSEN, M. T., JOHANSEN, J. V., CLOOS, P. A., RAPPSILBER, J. & HELIN, K. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature, 473, 343-8.

WINTER, J., JUNG, S., KELLER, S., GREGORY, R. I. & DIEDERICHS, S. 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat Cell Biol, 11, 228-34.

WISGERHOF, H. C., WOLTERBEEK, R., DE FIJTER, J. W., WILLEMZE, R. & BOUWES BAVINCK, J. N. 2012. Kidney Transplant Recipients with Cutaneous Squamous Cell Carcinoma Have an Increased Risk of Internal Malignancy. J Invest Dermatol, 132, 2176-2183.

WITKOS, T. M., KOSCIANSKA, E. & KRZYZOSIAK, W. J. 2011. Practical Aspects of microRNA Target Prediction. Curr Mol Med, 11, 93-109.

WITTEKINDT, C., WAGNER, S., MAYER, C. S. & KLUSSMANN, J. P. 2012. Basics of tumor development and importance of human papilloma virus (HPV) for head and neck cancer. GMS Curr Top Otorhinolaryngol Head Neck Surg, 11, Doc09.

WONG, C. S. M., STRANGE, R. C. & LEAR, J. T. 2003. Basal cell carcinoma. BMJ, 327, 794-798.

WOODCOCK, D. M., CROWTHER, P. J. & DIVER, W. P. 1987. The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide. Biochem Biophys Res Commun, 145, 888-94.

WU, H., D'ALESSIO, A. C., ITO, S., XIA, K., WANG, Z., CUI, K., ZHAO, K., SUN, Y. E. & ZHANG, Y. 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature, 473, 389-93.

WYSOCKI, A. B. 1999. Skin anatomy, physiology, and pathophysiology. Nurs Clin North Am, 34, 777-97, v.

XU, H. & ELMETS, C. A. 2012. UVB Immunosuppression: Vitamin D or Not Vitamin D[quest] That Is the Question. J Invest Dermatol, 132, 2676-2678.

XU, J., LI, X., YANG, H., CHANG, R., KONG, C. & YANG, L. 2013. SIN1 promotes invasion and metastasis of hepatocellular carcinoma by facilitating epithelial-mesenchymal transition. Cancer, 119, 2247-57.

XU, N., ZHANG, L., MEISGEN, F., HARADA, M., HEILBORN, J., HOMEY, B., GRANDER, D., STAHLE, M., SONKOLY, E. & PIVARCSI, A. 2012. MicroRNA-125b down-regulates matrix metallopeptidase 13 and inhibits cutaneous squamous cell carcinoma cell proliferation, migration, and invasion. J Biol Chem, 287, 29899-908.

XU, Y., WU, F., TAN, L., KONG, L., XIONG, L., DENG, J., BARBERA, A. J., ZHENG, L., ZHANG, H., HUANG, S., MIN, J., NICHOLSON, T., CHEN, T., XU, G., SHI, Y., ZHANG, K. & SHI, Y. G. 2011. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Molecular cell, 42, 451-64.

YACHIDA, S., JONES, S., BOZIC, I., ANTAL, T., LEARY, R., FU, B., KAMIYAMA, M., HRUBAN, R. H., ESHLEMAN, J. R., NOWAK, M. A., VELCULESCU, V. E., KINZLER, K. W., VOGELSTEIN, B. & IACOBUZIO-DONAHUE, C. A. 2010. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature, 467, 1114-7.

YAMAMOTO, H., KITADAI, Y., OUE, N., OHDAN, H., YASUI, W. & KIKUCHI, A. 2009. Laminin gamma2 mediates Wnt5a-induced invasion of gastric cancer cells. Gastroenterology, 137, 242-52, 252 e1-6.

YAMANE, K., JINNIN, M., ETOH, T., KOBAYASHI, Y., SHIMOZONO, N., FUKUSHIMA, S., MASUGUCHI, S., MARUO, K., INOUE, Y., ISHIHARA, T., AOI, J., OIKE, Y. & IHN, H. 2013. Down-regulation of miR-124/-214 in cutaneous squamous cell carcinoma mediates abnormal cell proliferation via the induction of ERK. J Mol Med (Berl), 91, 69-81.

YAMASHITA, J., IWAKIRI, T., FUKUSHIMA, S., JINNIN, M., MIYASHITA, A., HAMASAKI, T., MAKINO, T., AOI, J., MASUGUCHI, S., INOUE, Y. & IHN, H. 2013. The rs2910164 G>C polymorphism in microRNA-146a is associated with the incidence of malignant melanoma. Melanoma Res, 23, 13-20.

YAN, J., ZIERATH, J. R. & BARRÈS, R. 2011. Evidence for non-CpG methylation in mammals. Experimental cell research, 317, 2555-2561.

YAN, Z. H., BAO, Z. S., YAN, W., LIU, Y. W., ZHANG, C. B., WANG, H. J., FENG, Y., WANG, Y. Z., ZHANG, W., YOU, G., ZHANG, Q. G. & JIANG, T. 2013. Upregulation of DLX2 confers a poor prognosis in glioblastoma patients by inducing a proliferative phenotype. Curr Mol Med, 13, 438-45.

YANG, A., SCHWEITZER, R., SUN, D., KAGHAD, M., WALKER, N., BRONSON, R. T., TABIN, C., SHARPE, A., CAPUT, D., CRUM, C. & MCKEON, F. 1999. p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature, 398, 714-8.

YANG, X., LI, H., DENG, A. & LIU, X. 2010. Plk1 phosphorylation of Topors is involved in its degradation. Mol Biol Rep, 37, 3023-8.

YEGNASUBRAMANIAN, S., WU, Z., HAFFNER, M. C., ESOPI, D., ARYEE, M. J., BADRINATH, R., HE, T. L., MORGAN, J. D., CARVALHO, B., ZHENG, Q., DE MARZO, A. M., IRIZARRY, R. A. & NELSON, W. G. 2011. Chromosome-wide mapping of DNA methylation patterns in normal and malignant prostate cells reveals pervasive methylation of gene-associated and conserved intergenic sequences. BMC Genomics, 12, 313.

YI, R., POY, M. N., STOFFEL, M. & FUCHS, E. 2008. A skin microRNA promotes differentiation by repressing 'stemness'. Nature, 452, 225-9.

YIP, L., KELLY, L., SHUAI, Y., ARMSTRONG, M. J., NIKIFOROV, Y. E., CARTY, S. E. & NIKIFOROVA, M. N. 2011. MicroRNA signature distinguishes the degree of aggressiveness of papillary thyroid carcinoma. Ann Surg Oncol, 18, 2035-41.

YOOL, A. & EDMUNDS, W. J. 1998. Epigenetic inheritance and prions. Journal of Evolutionary Biology, 11, 241-242.

YOSHIMURA, K., AOKI, H., IKEDA, Y., FUJII, K., AKIYAMA, N., FURUTANI, A., HOSHII, Y., TANAKA, N., RICCI, R., ISHIHARA, T., ESATO, K., HAMANO, K. & MATSUZAKI, M. 2005. Regression of abdominal aortic aneurysm by inhibition of c-Jun N-terminal kinase. Nat Med, 11, 1330-8.

YOUSEFI, P., HUEN, K., SCHALL, R. A., DECKER, A., ELBOUDWAREJ, E., QUACH, H., BARCELLOS, L. & HOLLAND, N. 2013. Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. Epigenetics, 8, 1141-52.

YU, J., JUN, E., BAE, Y. & JUNG, J. 2008. Mesenchymal stem cells derived from human adipose tissues favour tumor cell growth in vivo. Stem Cells Dev, 17, 463 - 73.

YU, J., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, II & THOMSON, J. A. 2007. Induced pluripotent stem cell lines derived from human somatic cells. Science, 318, 1917-20.

YUAN, C. H., FILIPPOVA, M. & DUERKSEN-HUGHES, P. 2012. Modulation of apoptotic pathways by human papillomaviruses (HPV): mechanisms and implications for therapy. Viruses, 4, 3831-50.

ZACK, T. I., SCHUMACHER, S. E., CARTER, S. L., CHERNIACK, A. D., SAKSENA, G., TABAK, B., LAWRENCE, M. S., ZHANG, C.-Z., WALA, J., MERMEL, C. H., SOUGNEZ, C., GABRIEL, S. B., HERNANDEZ, B., SHEN, H., LAIRD, P. W., GETZ, G., MEYERSON, M. & BEROUKHIM, R. 2013. Pan-cancer patterns of somatic copy number alteration. Nat Genet, 45, 1134-1140.

ZAFIROPOULOS, A., TSENTELIEROU, E., BILLIRI, K. & SPANDIDOS, D. A. 2003. Human herpes viruses in non-melanoma skin cancers. Cancer Lett, 198, 77-81.

ZAIN, J., KAMINETZKY, D. & O'CONNOR, O. A. 2010. Emerging role of epigenetic therapies in cutaneous T-cell lymphomas. Expert Rev Hematol, 3, 187-203.

ZANDBERG, D. P., BHARGAVA, R., BADIN, S. & CULLEN, K. J. 2013. The role of human papillomavirus in nongenital cancers. CA Cancer J Clin, 63, 57-81.

ZAVADIL, J., CERMAK, L., SOTO-NIEVES, N. & BOTTINGER, E. P. 2004. Integration of TGF-[beta]/Smad and Jagged1/Notch signalling in epithelial-to-mesenchymal transition. EMBO J, 23, 1155-1165.

ZHANG, J., LI, S., YAN, Q., CHEN, X., YANG, Y., LIU, X. & WAN, X. 2013. Interferon-beta induced microRNA-129-5p down-regulates HPV-18 E6 and E7 viral gene expression by targeting SP1 in cervical cancer cells. PLoS One, 8, e81366.

ZHANG, J., NIU, C., YE, L., HUANG, H., HE, X., TONG, W. G., ROSS, J., HAUG, J., JOHNSON, T., FENG, J. Q., HARRIS, S., WIEDEMANN, L. M., MISHINA, Y. & LI, L. 2003. Identification of the haematopoietic stem cell niche and control of the niche size. Nature, 425, 836-41.

ZHANG, L., HUANG, J., YANG, N., GRESHOCK, J., MEGRAW, M. S., GIANNAKAKIS, A., LIANG, S., NAYLOR, T. L., BARCHETTI, A., WARD, M. R., YAO, G., MEDINA, A., O'BRIEN-JENKINS, A., KATSAROS, D., HATZIGEORGIOU, A., GIMOTTY, P. A., WEBER, B. L. & COUKOS, G. 2006a. microRNAs exhibit high frequency genomic alterations in human cancer. Proc Natl Acad Sci U S A, 103, 9136-41.

ZHANG, M. Q. 1999. Large-scale gene expression data analysis: a new challenge to computational biologists. Genome Res, 9, 681-8.

ZHANG, M. W., JIN, M. J., YU, Y. X., ZHANG, S. C., LIU, B., JIANG, X., PAN, Y. F., LI, Q. I., MA, S. Y. & CHEN, K. 2012. Associations of lifestyle-related factors, hsa-miR-149 and hsa-miR-605 gene polymorphisms with gastrointestinal cancer risk. Mol Carcinog, 51 Suppl 1, E21-31.

ZHANG, P., KERKELA, E., SKOTTMAN, H., LEVKOV, L., KIVINEN, K., LAHESMAA, R., HOVATTA, O. & KERE, J. 2007. Distinct sets of developmentally regulated genes that are expressed by human oocytes and human embryonic stem cells. Fertil Steril, 87, 677-90.

ZHANG, Y., YU, G., JIANG, P., XIANG, Y., LI, W., LEE, W. & ZHANG, Y. 2011. Decreased expression of protease-activated receptor 4 in human gastric cancer. Int J Biochem Cell Biol, 43, 1277-83.

ZHANG, Z., HUETTNER, P., NGUYEN, L., BIDDER, M., FUNK, M., LI, J. & RADER, J. 2006b. Aberrant promoter methylation and silencing of the POU2F3 gene in cervical cancer. Oncogene, 25, 5436 - 5445.

ZHAO, L. R., TIAN, W., WANG, G. W., CHEN, K. X. & YANG, J. L. 2013a. The prognostic role of PRUNE2 in leiomyosarcoma. Chin J Cancer, 32, 648-52.

ZHAO, Y., FANG, X., WANG, Y., ZHANG, J., JIANG, S., LIU, Z., MA, Z., XU, L., LI, E. & ZHANG, K. 2013b. Comprehensive Analysis for Histone Acetylation of Human Colon Cancer Cells Treated with a novel HDAC Inhibitor. Curr Pharm Des.

*ZHENG, Y., JOSEFOWICZ, S. Z., KAS, A., CHU, T. T., GAVIN, M. A. & RUDENSKY, A. Y. 2007. Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. Nature, 445***,** *936-40.*

*ZILLER, M. J., MULLER, F., LIAO, J., ZHANG, Y., GU, H., BOCK, C., BOYLE, P., EPSTEIN, C. B., BERNSTEIN, B. E., LENGAUER, T., GNIRKE, A. & MEISSNER, A. 2011. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet, 7***,** *e1002389.*